1932

9928-4

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

EDITED BY

EINAR HILLE S. LEFSCHETZ

WITH THE COÖPERATION OF THE

DEPARTMENT OF MATHEMATICS OF PRINCETON UNIVERSITY

AND OF

ORMOND STONE J. F. RITT
H. BATEMAN J. D. TAMARKIN
G. D. BIRKHOFF H. S. VANDIVER
OYSTEIN ORE A. PELL-WHEELER
NORBERT WIENER

PUBLISHED BY THE

PRINCETON UNIVERSITY PRESS

SECOND SERIES, VOL. 33

PRINCETON, N. J.

1932

LB
7A61a
v 33
1932
173368

LÜTCKE & WULFF, HAMBURG, GERMANY

TAUBERIAN THEOREMS.*

By Norbert Wiener.

		Page
INTRODUCTION		1
CHAPTER I.	The Closure of the Set of Translations of a Given Function	7
	1. Closure in class L_2	7
	2. Closure in class L_1	9
	3. A sub-class of L_1	21
CHAPTER II.	Asymptotic Properties of Averages	25
	4. Averages of bounded functions	25
	5. Averages of bounded Stieltjes distributions	26
	6. Averages of unilaterally bounded distributions and functions	28
CHAPTER III.	Tauberian Theorems and the Convergence of Series and Integrals	32
	7. The Hardy-Littlewood condition	32
	8. The Schmidt condition	36
CHAPTER IV.	Tauberian Theorems and Prime Number Theory	39
	9. Tauberian theorems and Lambert series	39
	10. Ikehara's theorem	44
CHAPTER V.	Special Applications of Tauberian Theorems	50
	11. On the proof of special Tauberian theorems	50
	12. Examples of kernels for which Tauberian theorems hold	52
	13. A theorem of Ramanujan	53
	14. The summation of trigonometrical developments	55
	15. Young's criterion for the convergence of a Fourier series	57
	16. Tauberian theorems and asymptotic series	60
CHAPTER VI.	Kernels Almost of the Closed Cycle	62
	17. The reduction of kernels almost of the closed cycle to kernels of the closed cycle	62
	18. A Tauberian theorem of Hardy and Littlewood	64
	19. The Tauberian theorem of Borel summation	67
CHAPTER VII.	A Quasi-Tauberian Theorem	72
	20. The quasi-Tauberian theorem	72
	21. Applications of the quasi-Tauberian theorem	77
CHAPTER VIII.	Tauberian Theorems and Spectra	79
	22. A further type of asymptotic behavior	79
	23. Generalized types of summability	90
	24. Some unsolved problems	93
BIBLIOGRAPHY		94

INTRODUCTION.

Numerous important branches of mathematics and physics concern themselves with the asymptotic behavior of functions for very large or very

* Received July 20, 1931.

1

small values of their arguments, or of certain parameters. Statistical mechanics is that branch of mechanics which concerns itself, not with an individual dynamical system of a finite number of degrees of freedom, but with the asymptotic behavior of dynamical systems as the number of degrees of freedom increases without limit. The analytical theory of numbers is likewise concerned with the behavior of assemblages of whole numbers as the size of these assemblages increases. To this domain of ideas belong asymptotic series in analysis, and a whole order of concepts clustering about the operational calculus of Heaviside and the Fourier series and integrals.

The Fourier integral is of peculiar interest in the study of asymptotic problems. If $f(x)$ is a function of Lebesgue class L_2, by a theorem of Plancherel,[1] there is related to it another function $g(u)$, also of L_2, defined by the equation

$$g(u) = \underset{A\to\infty}{\text{l. i. m.}} \frac{1}{\sqrt{2\pi}} \int_{-A}^{A} f(x)\, e^{iux}\, dx \tag{0.01}$$

(where l. i. m. stands for "limit in the mean"), such that

$$\int_{-\infty}^{\infty} |g(u)|^2\, du = \int_{-\infty}^{\infty} |f(x)|^2\, dx, \tag{0.02}$$

and

$$f(x) = \underset{A\to\infty}{\text{l. i. m.}} \frac{1}{\sqrt{2\pi}} \int_{-A}^{A} g(u)\, e^{-iux}\, du. \tag{0.03}$$

The functions $f(x)$ and $g(u)$ have the reciprocal relation, that asymptotic properties of each correspond to local properties of the other. They are known as Fourier transforms of one another. It is easy to see that $f(x+\lambda)$ and $g(u)\, e^{-iu\lambda}$ are likewise Fourier transforms of one another.

There is a large class of asymptotic problems in which the asymptotic property to be investigated is connected in some obvious and simple way with the entire class of functions $f(x+\lambda)$. Since the Fourier transforms of the functions of this class only differ by factors $e^{-iu\lambda}$, independent of the particular function $f(x)$, Fourier transformation is here a peculiarly useful tool. An example of such a problem is the investigation of the asymptotic behavior of the integral

$$\int_{-\infty}^{\infty} k(\lambda)\, f(x+\lambda)\, d\lambda. \tag{0.04}$$

Many problems which on first investigation do not appear to be concerned with integrals of the above sort may be put into such a form by an elementary transformation of variables. For example, if we put

[1] Cf. Plancherel (1), Titchmarsh (3), Mellin (1), Wiener (7). Cf. Bibliography on p. 94.

(0.05) $x = \log \xi, \quad \lambda = \log \mu, \quad f(x) = \varphi(\xi), \quad \mu k(\lambda) = Q(\mu),$

the integral in question assumes the form

(0.06) $$\int_0^\infty Q(\mu)\, \varphi(\xi\mu)\, d\mu.$$

Thus the study of the asymptotic properties of this integral also fall under those accessible through Fourier developments.

In 1925 a paper by Robert Schmidt[2] appeared in which the class of theorems known as Tauberian was brought into relation with the asymptotic properties of integrals of this type. Tauberian theorems gain their name from a theorem published by A. Tauber[3] in 1897, to the effect that if

(0.07) $$\lim_{x \to 1-0} \frac{1}{1-x} \sum_0^\infty a_n x^n = A,$$

and

(0.08) $$a_n = o(1/n),$$

then

(0.09) $$\sum_0^\infty a_n = A.$$

This is a conditioned converse of Abel's theorem, which stated that (0.07) follows from (0.09), without the mediation of any hypothesis such as (0.08). Such conditioned inverses of Abel's theorem, and of other analogous theorems which assert that the convergence of a series implies its summability by a certain method to the same sum, have been especially studied by G. H. Hardy and J. E. Littlewood, and have been termed by them *Tauberian*.

It is the service of Hardy and Littlewood[4] to have replaced hypothesis (0.08) by hypotheses of the form

(0.10) $$a_n = O\left(\frac{1}{n}\right),$$

or even of the form

(0.11) $$n a_n > -K.$$

The importance of these generalizations is scarcely to be exaggerated. They far exceed in significance Tauber's original theorem. The work of Hardy and Littlewood, unlike that of Tauber, makes very appreciable demands on analytical technique, and is capable, among other things, of supplying the gaps in Poisson's imperfect discussion of the convergence

[2] Schmidt (2).

[3] Tauber (1).

[4] Hardy (2), Littlewood (1), Hardy and Littlewood (4), (10), (13), (18) etc.

of the Fourier series. For these reasons, I feel that it would be far more appropriate to term these theorems Hardy-Littlewood theorems, were it not that usage has sanctioned the other appelation.

As we said, Tauberian theorems reduce to theorems on the conditioned equivalence of the asymptotic values of certain averages. The auxiliary conditions of Hardy and Littlewood, together with certain more extended conditions given by Schmidt[5] and discussed by Vijayaraghavan[6] and Szász,[7] become restrictions on the magnitude of the function or mass distribution of which the average is to be taken. An even further reduction eliminates this function from all consideration, and finds the essence of the Tauberian theorem in the linear properties of the weighting functions used in determining the average in question. The linear properties which are of importance concern the closure of the set of functions derivable from a given weighting function by translation, and the first chapter of this monograph is devoted to the study of such closure properties. The second chapter is occupied with the formulation and proof of the fundamental Tauberian theorems concerning averages, and the third with the transformation of these theorems into a recognized Tauberian form.

Some of the most interesting applications of Tauberian theorems have been in the analytic theory of numbers. Here two different avenues of approach are possible. Hardy and Littlewood[8] have shown that the prime number theorem, to the effect that the number of primes less than n is asymptotically $n/\log n$, is entirely equivalent to the Tauberian theorem concerning Lambert series, that if

$$\lim_{x\to 1-0} \sum_0^\infty \frac{n\, a_n\, x^n}{1-x^n} = A \tag{0.12}$$

and

$$n\, a_n > -K, \tag{0.11}$$

then

$$\sum_0^\infty a_n = A. \tag{0.09}$$

This theorem falls under the category of theorems demonstrable by the methods of the present paper, and involves no further information concerning the Riemann zeta function $\zeta(\sigma+\tau i)$ than that this must be free of zeros on the line $\sigma = 1$. This theorem was already contained in the author's first note on his method, dated 1928, and establishes the usefulness of Lambert series in the proof of the prime number theorem, which had

[5] Schmidt (1) and (2).

[6] Vijayaraghavan (1) and (2).

[7] Szász (1) and (2).

[8] Hardy and Littlewood (11).

frequently been questioned[9]. The proof was however needlessly complicated by the fact that after a good deal of labor was spent in transforming the author's average theorem into recognizable Tauberian form, this form did not appear to be especially direct in its relation to the prime number theorem. The present memoir gives a much more direct proof, starting from a theorem of the average type.

The other approach to the prime number theorem employs a theorem of S. Ikehara[10], which is itself a generalization of a theorem of Landau. Landau[11] showed that if $\sum a_n n^{-x}$ is a Dirichlet series with positive coefficients, representing a function $\psi(x)$ analytic on and to the right of $\Re(x) = 1$, except for a pole of order 1 at $x = 1$, for which the residue is A, and if the function in question is $O(|x|^k)$ for $\Re(x) \geqq 1$, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_0^n a_l = A. \tag{0.13}$$

Ikehara proved by a Tauberian theorem that the condition that $\psi(x) = O(|x|^k)$ is inessential. By applying the theorem to $\psi(x) = -\zeta'(x)/\zeta(x)$ it is at once seen that the prime number theorem again follows from the fact that the Riemann zeta function has no zeros on the line $\Re(x) = 1$.

Chapter V is devoted to miscellaneous applications of Tauberian theorems. Among these perhaps the most important are to trigonometric developments and their summability. Here a theorem of Ramanujan[12] on Fourier transforms is of service. A particular application of Tauberian theorems to the criterion of Young[13] for the convergence of a Fourier series is due to Dr. Littauer[14], and is here discussed. We also take up the application of Tauberian theorems to asymptotic series and Wintner's[15] work.

There is a further field of application for Tauberian theorems where we need to introduce something like Robert Schmidt's[16] notion of "gestrahlte Matrizen". This we discuss in Chapter VI. In this category we find certain theorems of Hardy and Littlewood[17] related to Abel summation, as also the Tauberian theorems that concern themselves with Borel summation[18].

[9] Wiener (4).

[10] Ikehara (1).

[11] Landau (3).

[12] Hardy and Littlewood (8).

[13] Young (1).

[14] Littauer (2).

[15] Wintner (1).

[16] Schmidt (2).

[17] Cf. Hardy and Littlewood (10), (13), (18); Karamata (1), (2), (3), (4); Szász (1), (2); Doetsch (3); etc.

[18] Cf. Hardy and Littlewood (2), (3); Doetsch (4); Schmidt (1); Vijayaraghavan (1); Wiener (4).

In Chapter VII, we discuss certain theorems not strictly of a Tauberian nature, in that they involve no positiveness or boundedness condition. We here follow an earlier paper[19] in which the author proved the Hardy-Littlewood necessary and sufficient condition for the summability of a Fourier series. There our result was not a "best possible" one, as the work of Bosanquet[20] and Paley has demonstrated. At present, our method has succeeded in yielding their strict theorem.

Chapter VIII is devoted to the development of certain Tauberian theorems intimately connected with our generalized harmonic analysis[21]. These suggest certain new definitions of summability, and we discuss definitions of this sort.

The genesis of the present paper may deserve a word of comment. In the preparation of a previous investigation on generalized harmonic analysis, the author found himself obliged to make use of certain theorems communicated to him by Mr. A. E. Ingham and of a Tauberian nature. A correspondence arose with Professor Hardy, finally resulting in a number of papers of a Tauberian character. While interested in theorems of this type, the author had the work of Dr. Robert Schmidt brought to his notice, and at Dr. Schmidt's suggestion, began to search for a method of combining the trigonometric attack of his own earlier papers with the generality of the Schmidt standpoint. It was only by a radical change in the manner in which trigonometric methods were applied that this attempt succeeded. This change consisted in a logarithmic change of base before the introduction of the harmonic analysis. It was quite along lines contemplated by Schmidt, who correlated with his Tauberian theorems a certain moment problem. Schmidt, however, found in the problem of uniqueness the significant aspect of this moment problem. The present author looked for it rather in the problem of the existence of a solution together with certain associated problems of approximation. It was Dr. Schmidt himself who furnished the experimentum crucis which established the greater scope of the methods of this paper. He suggested that the Tauberian theorem for Lambert series had resisted his methods of proof, and that it might be desirable to try on it the edge of any new method.

The origination of a new method is but the prelude to a large amount of detailed application before its power can be judged or its limitations defined. In this task the author has had the valuable aid of his students, Drs. S. B. Littauer and S. Ikehara. He also wishes to express his gratitude for the criticisms of Professors Hardy and Tamarkin at more than one stage of his work.

[19] Cf. Hardy and Littlewood (12); Wiener (5).

[20] Bosanquet (1); Paley (1).

[21] Wiener (7).

CHAPTER I.

THE CLOSURE OF THE SET OF TRANSLATIONS OF A GIVEN FUNCTION.

1. **Closure in class L_2.** Let $f(x)$ be a function, real or complex, defined for all real arguments over $(-\infty, \infty)$. Let it belong to Lebesgue class L_2—that is, let it be measurable, and let

$$\int_{-\infty}^{\infty} |f(x)|^2\, dx \tag{1.01}$$

be finite. We shall term the class of all functions $f(x+\lambda)$ for all real values of λ the class of *translations of* $f(x)$. We wish to know when this class is closed or complete—that is, when it is possible, whenever a function $F(x)$ from L_2 is given, and any positive quantity ε, to find a function $F_1(x)$ of the form

$$F_1(x) = \sum_1^n A_k f(x+\lambda_k), \tag{1.02}$$

such that

$$\int_{-\infty}^{\infty} |F(x) - F_1(x)|^2\, dx \leqq \varepsilon. \tag{1.03}$$

The situation is governed by the following theorem:

THEOREM I. *The necessary and sufficient condition for the set of all translations of $f(x)$ to be closed is that the real zeros of its Fourier transform*

$$g(u) = \text{l. i. m.}_{A\to\infty} \frac{1}{\sqrt{2\pi}} \int_{-A}^{A} f(x)\, e^{iux}\, dx \tag{0.01}$$

should form a set of zero measure.

The existence of this Fourier transform results from a familiar theorem of Plancherel.[1] This theorem further requires that

$$\int_{-\infty}^{\infty} |g(u)|^2\, du = \int_{-\infty}^{\infty} |f(x)|^2\, dx \tag{0.02}$$

and that

$$f(x) = \text{l. i. m.}_{A\to\infty} \frac{1}{\sqrt{2\pi}} \int_{-A}^{A} g(u)\, e^{-iux}\, du. \tag{0.03}$$

The Fourier transform of $f(x+\lambda)$ is $e^{-iu\lambda} g(u)$.

On account of (0.02), properties of convergence in the mean are conserved under a Fourier transformation, for these concern merely the integral of the square of the modulus of the difference of two functions, which is equal to the integral of the square of the difference of their Fourier transforms. Accordingly, Theorem I is a corollary of:

LEMMA Ia. *If $g(u)$ is a function belonging to L_2, then the set of functions $e^{-iu\lambda}g(u)$ is closed when and only when the zeros of $g(u)$ form a set of zero measure.*

Since every zero of $g(u)$ is also a zero of $e^{-iu\lambda}g(u)$, the necessity of this condition is at once obvious, for if

$$G_1(u) = \sum_1^n A_k e^{-iu\lambda_k} g(u) \tag{1.04}$$

is any linear combination of the functions $e^{-iu\lambda_k}g(u)$, then

$$\int_A^B |G_1(u) - 1|^2 du \tag{1.05}$$

must equal or exceed the measure of the set of the points between A and B where $g(u)$ vanishes. That is, if the measure of this set is not zero, the function which is 1 between A and B and zero elsewhere cannot be a limit in the mean of functions $G_1(u)$.

Now as to the sufficiency of the condition in Lemma Ia: let

$$\varphi(u) = \begin{cases} 1 - \dfrac{|u|}{A}, & [|u| \leqq A]; \\ 0, & [|u| > A]. \end{cases} \tag{1.06}$$

Then

$$\varphi(u) = \frac{1}{\pi A}\int_{-\infty}^{\infty} e^{-iu\mu}\frac{1-\cos A\mu}{\mu^2}\,d\mu. \tag{1.07}$$

This is an absolutely convergent integral, and hence a limit of a sequence of polynomials

$$\sum A_n e^{-iu\lambda_n} \tag{1.08}$$

converging boundedly, and uniformly over every finite interval. Hence

$$e^{-i\frac{un\pi}{A}}\varphi(u)g(u) \tag{1.09}$$

is a limit in the mean of such functions as those in (1.04).

Now let $\psi(u)$ be a quadratically summable function satisfying the conditions

$$\int_{-A}^{A}\psi(u)\varphi(u)g(u)e^{\frac{-iun\pi}{A}}du = 0. \tag{1.10}$$

$$[n = \cdots, -2, -1, 0, 1, 2, \cdots]$$

Since $\psi(u)\varphi(u)g(u)$ is absolutely summable,[22] we shall have

$$\psi(u)\varphi(u)g(u) = 0 \tag{1.11}$$

[22] This follows from the Parseval theorem for Fourier integrals.

except over a set of measure zero.[23] If $g(u)$ does not vanish except over a set of measure zero, since $\varphi(u)$ has no zeros, $\psi(u)$ vanishes over $(-A, A)$ except over a set of measure zero. Thus if $g(u) \not\equiv 0$, there is no quadratically summable function not almost everywhere zero, orthogonal to every function $e^{\frac{-iun\pi}{A}}\varphi(u)g(u)$ over $(-A, A)$. This means that every function of L_2 vanishing everywhere outside $(-A, A)$ is a limit in the mean of polynomials[24] in $e^{\frac{-iun\pi}{A}}\varphi(u)g(u)$ and hence in $e^{-iu\lambda_n}g(u)$. Every function of L_2 is however the limit in the mean of a sequence of functions of L_2 vanishing for large arguments. Thus Lemma Ia and Theorem I are established.

2. **Closure in class L_1.** The Lebesgue class L_1 consists of all measurable functions $f(x)$ which are absolutely integrable over $(-\infty, \infty)$. If $f(x)$ is such a function, its Fourier transform

$$g(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{iux}\, dx \tag{2.01}$$

of course exists, and is bounded and continuous. As before, the Fourier transform of $f(x+\lambda)$ is $e^{-iu\lambda}g(u)$.

We shall say that a class C of functions $\varphi_\lambda(x)$ of L_1 is closed L_1 if, whenever $F(x)$ is a function of L_1 and ϵ is a positive number, there is a polynomial

$$F_1(x) = \sum_1^n A_k \varphi_{\lambda_k}(x) \tag{2.02}$$

of functions of C, such that

$$\int_{-\infty}^{\infty} |F(x) - F_1(x)|\, dx \leqq \epsilon. \tag{2.03}$$

We shall prove the following theorem:

THEOREM II. *If $f(x)$ is a function of L_1, a necessary and sufficient condition for the set of all translations of $f(x)$ to be closed L_1 is that its Fourier transform*

$$g(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{iux}\, dx \tag{2.01}$$

should have no real zeros.

---

[23] This follows from the Riemann theorem on the unicity of the function with a given Fourier series.

[24] Here we make use of the familiar theorem that if any linear class of functions of L_2 is given, every function of L_2 is the sum of a function expressible in the mean by the functions of the class and a function orthogonal to all functions of the class.

The necessity of this condition is again obvious. If $g(u_1) = 0$, the same is true for the argument u_1 of the Fourier transform of

$$F_1(x) = \sum_1^n A_k f(x+\lambda_k). \tag{1.02}$$

Let $f_1(x)$ be a function of L_1 with a Fourier transform $g_1(x)$ for which

$$g_1(u) \not\equiv 0.$$

Then

$$\int_{-\infty}^{\infty} |F_1(x)-f_1(x)|\, dx \geqq \left| \int_{-\infty}^{\infty} (f_1(x)-F_1(x))\, e^{iu_1 x}\, dx \right| = \sqrt{2\pi}\, |g_1(u_1)| > 0. \tag{2.04}$$

This proves that f_1 is not a limit in the mean (L_1) of any sequence of F_1's.

To prove the sufficiency of the condition of Theorem II, we shall have recourse to a sequence of lemmas.

LEMMA IIa. *If*

$$f_1(x) = \sum_{-\infty}^{\infty} a_n\, e^{inx}, \qquad f_2(x) = \sum_{-\infty}^{\infty} b_n\, e^{inx}, \tag{2.05}$$

and

$$\sum_{-\infty}^{\infty} |a_n| = A < \infty, \qquad \sum_{-\infty}^{\infty} |b_n| = B < \infty, \tag{2.06}$$

and if

$$f_1(x)\, f_2(x) \sim \sum_{-\infty}^{\infty} c_n\, e^{inx} \tag{2.07}$$

then

$$\sum_{-\infty}^{\infty} |c_n| \leqq AB. \tag{2.08}$$

The proof is immediate, for

$$\sum_{-\infty}^{\infty} |c_n| = \sum_{n=-\infty}^{\infty} \left| \sum_{k=-\infty}^{\infty} a_k\, b_{n-k} \right| \leqq \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |a_k|\, |b_{n-k}| = \sum_{-\infty}^{\infty} |a_n| \sum_{-\infty}^{\infty} |b_n| = AB. \tag{2.09}$$

LEMMA IIb. *If $f(x)$ is a function of period 2π, and if at every point y, there is a function $f_y(x)$, coincident with $f(x)$ over some interval $(y-\varepsilon_y, y+\varepsilon_y)$, of period 2π, and such that its Fourier series converges absolutely, then the Fourier series of $f(x)$ converges absolutely*[25].

[25] We shall say that a Fourier series converges absolutely when the sum of the moduli of the coefficients converges, the series being represented in complex exponential form. This is of course equivalent to the absolute convergence of the series in this form for any single real argument.

To prove this, let us reflect that by the Heine-Borel theorem, any period of $f(x)$ may be covered by a finite number of overlapping intervals $(y-\varepsilon_y,\ y+\varepsilon_y)$. Let such an interval be $(A,\ B)$, let its neighbor to the left have its right-hand end-point at C, and let its neighbor to the right have its left-hand end-point at D. There will be no loss in generality if we suppose that $A<C<D<B$. Let us define $\varphi_y(x)$ by

$$\varphi_y(x) = \begin{cases} 0 \quad ; & [-\pi \leqq x < A] \\ \dfrac{x-A}{c-A}\,; & [\ A \leqq x < C] \\ 1 \quad ; & [\ C \leqq x < D] \\ \dfrac{B-x}{B-D}\,; & [\ D \leqq x < B] \\ 0 \quad . & [\ B \leqq x < \pi]. \end{cases} \tag{2.10}$$

It is easy to show that the Fourier series of $\varphi_y(x)$ converges absolutely. Thus the Fourier series of

$$\varphi_y(x)\, f(x) = \varphi_y(x)\, f_y(x) \tag{2.11}$$

converges absolutely, by Lemma IIa. However, if we add together the functions $\varphi_y(x)$ for a set of intervals completely covering a period of $f(x)$, the sum will be $f(x)$ itself. Thus $f(x)$ will be the sum of a finite number of functions with absolutely convergent Fourier series and must itself have an absolutely convergent Fourier series.

LEMMA IIc. *If*

$$f(x) = \sum_{-\infty}^{\infty} a_n\, e^{inx} \tag{2.12}$$

and

$$|a_0| > \sum_{-\infty}^{-1} |a_n| + \sum_{1}^{\infty} |a_n|, \tag{2.13}$$

then the Fourier series of $1/f(x)$ *converges absolutely.*

We may write

$$\begin{aligned} \frac{1}{f(x)} &= \frac{1}{a_0 + \sum_{-\infty}^{-1} a_n\, e^{inx} + \sum_{1}^{\infty} a_n\, e^{inx}} \\ &= \frac{1}{a_0}\left\{1 - \frac{1}{a_0}\left[\sum_{-\infty}^{-1} a_n\, e^{inx} + \sum_{1}^{\infty} a_n\, e^{inx}\right]\right. \\ &\qquad \left. + \frac{1}{a_0^2}\left[\sum_{-\infty}^{-1} a_n\, e^{inx} + \sum_{1}^{\infty} a_n\, e^{inx}\right]^2 - \cdots\right\}. \end{aligned} \tag{2.14}$$

That this is more than formally true, and that $1/f(x)$ is continuous, results from the fact that the geometrical progression

$$(2.15)\quad \frac{1}{|a_0|}\left\{1+\frac{1}{|a_0|}\left[\sum_{-\infty}^{-1}|a_n|+\sum_{1}^{\infty}|a_n|\right]+\frac{1}{a_0^2}\left[\sum_{-\infty}^{-1}|a_n|+\sum_{1}^{\infty}|a_n|\right]^2+\cdots\right\}$$

converges. Term by term, series (2.15) is greater than or equal to the sum of the sums of the moduli of the coefficients of the Fourier series of the successive terms of (2.14), and hence is greater than or equal to the sum of the moduli of the coefficients of the Fourier series of $1/f(x)$.

LEMMA IId. *Let*

$$(2.16)\quad \sum_{-\infty}^{\infty}|a_n| = A<\infty,$$

and let

$$(2.17)\quad f(x) = \sum_{-\infty}^{\infty} a_n e^{inx}.$$

Let

$$f(y) \neq 0.$$

Then there is a neighborhood $(y-\varepsilon,\ y+\varepsilon)$ of y and a function $f_y(x)$ with absolutely convergent Fourier series such that over $(y-\varepsilon,\ y+\varepsilon)$,

$$(2.18)\quad f_y(x) = f(x)$$

and that the Fourier series of $f_y(x)$, let us say

$$(2.19)\quad \sum_{-\infty}^{\infty} c_n e^{inx}$$

has the property that

$$(2.20)\quad |c_0| > \sum_{-\infty}^{-1}|c_n| + \sum_{1}^{\infty}|c_n|.$$

There is manifestly no restriction in taking y to be 0. Let us introduce the auxiliary function

$$(2.201)\quad \varphi_\varepsilon(x) = \begin{cases} 1 & ;\ [|x|<\varepsilon] \\ 2-\dfrac{|x|}{\varepsilon};\ [\varepsilon\leqq|x|<2\varepsilon] \\ 0 & .\ [2\varepsilon\leqq|x|]. \end{cases}$$

This will have the Fourier series

$$(2.202)\quad \frac{3\varepsilon}{2\pi}+\sum_{1}^{\infty}\frac{e^{inx}+e^{-inx}}{\pi n^2\varepsilon}(\cos\varepsilon n-\cos 2\varepsilon n).$$

If we represent the function $\{f(x)\varphi_\varepsilon(x)+f(0)[1-\varphi_\varepsilon(x)]\}$ by the Fourier series

$$\sum_{-\infty}^{\infty} c_n e^{inx},$$

we shall have

$$c_n = \frac{3\varepsilon}{2\pi} a_n + \sum_{k=1}^{\infty} \frac{a_{n-k} + a_{n+k}}{\pi k^2 \varepsilon} (\cos \varepsilon k - \cos 2\varepsilon k)$$

$$- \sum_{-\infty}^{\infty} a_k \frac{\cos \varepsilon n - \cos 2\varepsilon n}{\pi n^2 \varepsilon}; \qquad [n \neq 0],$$

(2.203)

$$c_0 = a_0 + \sum_{1}^{\infty} (a_{-k} + a_k) \left[1 + \frac{\cos \varepsilon k - \cos 2\varepsilon k}{\pi k^2 \varepsilon} - \frac{3\varepsilon}{2\pi} \right].$$

Hence

$$\lim_{\varepsilon \to 0} |c_0| = \left| \sum_{-\infty}^{\infty} a_k \right| = |f(0)| > 0, \tag{2.204}$$

and

$$\sum_{1}^{\infty} [|c_n| + |c_{-n}|]$$

$$\leqq \sum_{m=-\infty}^{\infty} |a_m| \left\{ \sum_{n=-\infty}^{\infty}{}'' \left| \frac{\cos \varepsilon n - \cos 2\varepsilon n}{\pi n^2 \varepsilon} - \frac{\cos \varepsilon (m-n) - \cos 2\varepsilon (m-n)}{\pi (m-n)^2 \varepsilon} \right| \right.$$

$$\left. + \left| \frac{\cos \varepsilon m - \cos 2\varepsilon m}{\pi m^2 \varepsilon} - \frac{3\varepsilon}{2\pi} \right| \right\} \tag{2.205}$$

$$= \sum_{-\infty}^{\infty} |a_m| A_m.$$

Here $\sum''$ means that the values 0 and m are omitted.

We have

$$\left| \frac{\cos \varepsilon n - \cos 2\varepsilon n)}{\pi n^2 \varepsilon} \right| = \left| \frac{2 \sin \frac{\varepsilon n}{2} \sin \frac{3\varepsilon n}{2}}{\pi n^2 \varepsilon} \right| \leqq \begin{cases} \frac{3\varepsilon}{2\pi}; \\ \frac{2}{\pi n^2 \varepsilon}; \end{cases} \tag{2.206}$$

and

$$\left| \frac{\cos \varepsilon n - \cos 2\varepsilon n}{\pi n^2 \varepsilon} - \frac{\cos \varepsilon (m-n) - \cos 2\varepsilon (m-n)}{\pi (m-n)^2 \varepsilon} \right|$$

$$= \left| \int_{n-m}^{n} \frac{d}{dx} \left(\frac{\cos \varepsilon x - \cos 2\varepsilon x}{\pi x^2 \varepsilon} \right) dx \right| \tag{2.21}$$

$$\leqq \frac{\varepsilon^2}{\pi} \left| \int_{n-m}^{n} \left| \max_{y} \frac{d}{dy} \frac{\cos y - \cos 2y}{y^2} \right| dx \right| \leqq |m| \varepsilon^2 c.$$

This formula may be applied in the limiting form whenever $m = n$ or $m = 0$. It results from (2.205) and (2.206) that for all sufficiently small ε

$$A_m \leqq 4 \left[\frac{1}{\varepsilon} + 1 \right] \frac{3\varepsilon}{2\pi} + \frac{8}{\pi \varepsilon} \sum_{[1/\varepsilon]}^{\infty} n^{-2} < \frac{15}{\pi}, \tag{2.22}$$

and from (2.205) and (2.21) that for all sufficiently small ε

$$A_m \leqq 2[\varepsilon^{-3/2}+1]\,|m|\,\varepsilon^2 c + \frac{8}{\pi\varepsilon}\sum_{[\varepsilon^{-3/2}]}^{\infty} n^{-2} \tag{2.23}$$
$$< \varepsilon^{1/2}[2|m|c+9/\pi].$$

Combining (2.205), (2.22) and (2.23), we get

$$\lim_{\varepsilon\to 0}\sum_{-\infty}^{\infty}|a_m|\,A_m = 0, \tag{2.24}$$

which with (2.204) yields us

$$\sum_{1}^{\infty}[|c_n|+|c_{-n}|] < |c_0|. \tag{2.25}$$

If we take

$$f_\varepsilon(x) = f(x)\,\varphi_\varepsilon(x) + f(0)\,(1-\varphi_\varepsilon(x)), \tag{2.26}$$

Lemma IId is established.

As a direct consequence of Lemmas IIb, IIc, and IId we obtain:

LEMMA IIe. *If $f(x)$ is a function with an absolutely convergent Fourier series, which nowhere vanishes for real arguments, $1/f(x)$ has an absolutely convergent Fourier series.*

To prove this let us note that, in every neighborhood, by IId, $f(x)$ coincides with a function which by IIc has a reciprocal with an absolutely convergent Fourier series. By IIa, since in every neighborhood $1/f(x)$ coincides with a function with an absolutely convergent Fourier series, its Fourier series converges absolutely.

LEMMA IIf. *Let $f(x)$ be a function with an absolutely convergent Fourier series over the interval $(-\pi, \pi)$, and vanishing over neighborhoods including π and $-\pi$. Let $f(x)$ vanish everywhere outside $(-\pi, \pi)$. Let*

$$g(u) = \frac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi} f(x)\,e^{iux}\,dx. \tag{2.27}$$

Then the integral

$$\int_{-\infty}^{\infty}|g(u)|\,du \tag{2.28}$$

will converge. Conversely, if $f(x)$ vanishes over the region indicated, and (2.28) converges, its Fourier series will converge absolutely.

Let ε be such that $f(x)$ has no non-zero, values outside $(-\pi+2\varepsilon, \pi-2\varepsilon)$. Let us define $\varphi(x)$ by:

$$\varphi(x) = \begin{cases} 1 & ; \ [|x| < \pi - \varepsilon] \\ \dfrac{\pi - |x|}{\varepsilon} & ; \ [\pi - \varepsilon \leqq |x| < \pi] \\ 0 & . \ [\pi \leqq |x|] \end{cases} \tag{2.29}$$

Let

$$f(x) = \sum_{-\infty}^{\infty} a_n e^{inx}. \tag{2.17}$$

We shall have

$$\varphi(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} e^{-iux} \frac{\cos u(\pi - \varepsilon) - \cos u\pi}{u^2 \varepsilon} du, \tag{2.30}$$

and hence

$$\begin{aligned} f(x) = f(x)\varphi(x) &= \frac{1}{\pi} \int_{-\infty}^{\infty} du \sum_{-\infty}^{\infty} a_n e^{-i(u-n)x} \frac{\cos u(\pi - \varepsilon) - \cos u\pi}{u^2 \varepsilon} \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} e^{iux} \left[\sum_{-\infty}^{\infty} a_n \frac{\cos(u+n)(\pi - \varepsilon) - \cos(u+n)\pi}{(u+n)^2 \varepsilon} \right] du, \end{aligned} \tag{2.31}$$

where the rearrangement is possible since both integral and sum form part of an absolutely convergent double summation process. Then

$$g(u) = \sqrt{\frac{2}{\pi}} \sum_{-\infty}^{\infty} a_n \frac{\cos(u+n)(\pi - \varepsilon) - \cos(u+n)\pi}{(u+n)^2 \varepsilon} \tag{2.32}$$

will satisfy the condition that (2.28) shall be finite, and equation (2.27).

To prove the converse part of Lemma IIf, let us notice that as a result of the convergence of (2.28),

$$\sum_{-\infty}^{\infty} \left| \int_{n-1/2}^{n+1/2} g(u)\, du \right| \tag{2.33}$$

converges. Now,

$$\begin{aligned} \int_{n-1/2}^{n+1/2} g(u)\, du &= \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(x)\, dx \int_{n-1/2}^{n+1/2} e^{iux}\, du \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(x) \frac{2 \sin x/2}{x} e^{inx}\, dx. \end{aligned} \tag{2.34}$$

Thus the Fourier series of

$$f(x) \frac{2 \sin x/2}{x} \tag{2.35}$$

converges absolutely. Since the same is true of the Fourier series of

$$\frac{x}{2 \sin x/2} \varphi(x) \tag{2.36}$$

as may be determined by a simple computation, it follows from Lemma IIa that the same is true for the Fourier series of $f(x)$.

LEMMA IIg. *Let $f(x)$ be a function of class L_1. Then*

$$\lim_{\varepsilon \to 0} \int_{-\infty}^{\infty} |f(x+\varepsilon) - f(x)|\, dx = 0. \tag{2.37}$$

This is a well-known theorem and is proved in Hobson's *Theory of Functions of a Real Variable.*

LEMMA IIh. *Let $f(x)$ be a function of class L_1. Then if $\Phi(y)$ is any function of class L_1, differing from 0 only over $(-\varepsilon, \varepsilon)$,*

$$\int_{-\infty}^{\infty} \left| f(x) \int_{-\infty}^{\infty} \Phi(y)\, dy - \int_{-\infty}^{\infty} f(x+y)\, \Phi(y)\, dy \right| dx \tag{2.38}$$
$$\leqq \int_{-\infty}^{\infty} |\Phi(y)|\, dy \max_{|z| \leqq \varepsilon} \int_{-\infty}^{\infty} |f(x+z) - f(x)|\, dx.$$

The proof of this lemma merely depends on the inversion of the order of integration in an absolutely convergent double integral.

LEMMA IIi. *Let $f(x)$ be a function of class L_1. Then*

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} \left| f(x) - \frac{1}{\pi n} \int_{-\infty}^{\infty} f(x+y) \frac{\sin^2 n y}{y^2}\, dy \right| dx = 0. \tag{2.39}$$

This is a theorem of the Fejér type. It depends on the splitting of

$$\Phi_0(y) = \frac{\sin^2 n y}{y^2}$$

into the sum $\Phi_1(y) + \Phi_2(y)$ where

$$\Phi_1(y) = \begin{cases} \Phi_0(y)\,(1 - |y| n^{1/2}); & [|y| \leqq n^{-1/2}] \\ 0; & [|y| > n^{-1/2}]. \end{cases} \tag{2.40}$$

Lemmas IIg and IIh enable us to show that

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} \left| f(x) - \frac{1}{\pi n} \int_{-\infty}^{\infty} f(x+y)\, \Phi_1(y)\, dy \right| dx = 0 \tag{2.41}$$

while

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} \left| \frac{1}{\pi n} \int_{-\infty}^{\infty} f(x+y)\, \Phi_2(y)\, dy \right| dx \tag{2.42}$$
$$\leqq \int_{-\infty}^{\infty} |f(x)|\, dx \lim_{n \to \infty} \frac{1}{\pi n} \int_{-\infty}^{\infty} \Phi_2(y)\, dy.$$

Now

$$\frac{1}{\pi n}\int_{-\infty}^{\infty}\Phi_2(y)\,dy$$

$$=\frac{1}{\pi n}\int_{-\infty}^{\infty}\frac{\sin^2 ny}{y^2}\,dy-\frac{1}{\pi n}\int_{-n^{-1/2}}^{n^{-1/2}}(1-|y|n^{1/2})\frac{\sin^2 ny}{y^2}\,dy$$

$$(2.421)\qquad =\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin^2 z}{z^2}\,dz-\frac{1}{\pi}\int_{-n^{1/2}}^{n^{1/2}}(1-|z|n^{-1/2})\frac{\sin^2 z}{z^2}\,dz$$

$$=\frac{2}{\pi}\int_{n^{1/2}}^{\infty}\frac{\sin^2 z}{z^2}\,dz+\frac{2}{\pi n^{1/2}}\int_{0}^{n^{1/2}}\frac{\sin^2 z}{z}\,dz$$

$$=O(n^{-1/2}\log n).$$

Hence

$$(2.422)\qquad \lim_{n\to\infty}\int_{-\infty}^{\infty}\left|\frac{1}{\pi n}\int_{-\infty}^{\infty}f(x+y)\,\Phi_2(y)\,dy\right|dx=0,$$

and Lemma IIi is established.

We are now in a position to proceed to the proof of Theorem II. Let $F(x)$ be any function of L_1. By Lemma IIi, we can find a function $F_\eta(x)$ of the form

$$(2.43)\qquad \frac{1}{\pi n}\int_{-\infty}^{\infty}F(x+y)\,\frac{\sin^2 ny}{y^2}\,dy,$$

such that

$$(2.44)\qquad \int_{-\infty}^{\infty}|F(x)-F_\eta(x)|\,dx<\eta.$$

The Fourier transform $H_1(u)$ of this function $F_\eta(x)$ is

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{iux}\,dx\,\frac{1}{\pi n}\int_{-\infty}^{\infty}F(x+y)\,\frac{\sin^2 ny}{y^2}\,dy$$

$$(2.45)\qquad =\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}F(x)\,e^{iux}\,dx\,\frac{1}{\pi n}\int_{-\infty}^{\infty}\frac{\sin^2 ny}{y^2}\,e^{-iuy}\,dy$$

$$=\begin{cases}\left(1-\dfrac{|u|}{2n}\right)\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{\infty}F(x)\,e^{iux}\,dx; & [|u|<2n]\\ 0 \qquad\qquad ; & [|u|\geqq 2n]\end{cases}$$

and will vanish for all values of its argument u larger than $2n$ in modulus. The same will be the case with the Fourier transform $H_2(u)$ of

$$(2.46)\qquad \frac{1}{2\pi n}\int_{-\infty}^{\infty}f(x+y)\,\frac{\sin^2 2ny}{y^2}\,dy$$

which will vanish for values of its argument greater in modulus than $4n$, and only for such values. If we expand these Fourier transforms in Fourier series over $(-8n, 8n)$, by Lemma IIf, these series converge absolutely.

By an argument substantially identical with that used in proving Lemma IIb, we see that in the neighborhood of every point of $(-3n, 3n)$, there is a function coinciding locally with $1/H_2(u)$, and with an absolutely convergent Fourier series. It follows from Lemma IIa that the same is true of $H_1(u)/H_2(u)$ over the same interval. Since 0 enjoys this property also, the function which is $H_1(u)/H_2(u)$ over $(-3n, 3n)$ and 0 over the rest of a period $(-8n, 8n)$ has an absolutely convergent Fourier series, and hence, by Lemma IIf, an absolutely convergent Fourier integral. That is, we may write

$$H_1(u) = H_2(u)\, H_3(u) \tag{2.47}$$

where $H_3(u)$ is of the form

$$H_3(u) = \int_{-\infty}^{\infty} \Phi(x)\, e^{iux}\, dx \tag{2.48}$$

and

$$\int_{-\infty}^{\infty} |\Phi(x)|\, dx \tag{2.49}$$

converges. We shall then have

$$\begin{aligned} &\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F_\eta(x)\, e^{iux}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{iux}}{2\pi n}\, dx \int_{-\infty}^{\infty} f(x+y)\, \frac{\sin^2 2ny}{y^2}\, dy \int_{-\infty}^{\infty} \Phi(z)\, e^{iuz}\, dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{iux}\, dx \\ &\quad \times \left[\frac{1}{2\pi n} \int_{-\infty}^{\infty} f(x+\omega)\, d\omega \int_{-\infty}^{\infty} \frac{\sin^2 2ny}{y^2}\, \Phi(y-\omega)\, dy \right], \end{aligned} \tag{2.50}$$

or

$$\begin{aligned} &\int_{-\infty}^{\infty} e^{iux}\, dx \\ &\times \left[F_\eta(x) - \frac{1}{2\pi n} \int_{-\infty}^{\infty} f(x+\omega)\, d\omega \int_{-\infty}^{\infty} \frac{\sin^2 2ny}{y^2}\, \Phi(y-\omega)\, dy \right] = 0. \end{aligned} \tag{2.51}$$

Here we may replace e^{iux} by any function with an absolutely convergent Fourier integral, and hence by functions positive over an arbitrarily small region and zero everywhere else. This can only be if

$$F_\eta(x) = \frac{1}{2\pi n}\int_{-\infty}^{\infty} f(x+\omega)\,d\omega \int_{-\infty}^{\infty} \frac{\sin^2 2ny}{y^2}\,\Phi(y-\omega)\,dy$$
$$= \int_{-\infty}^{\infty} f(x+\omega)\,\Psi(\omega)\,d(\omega). \tag{2.52}$$

Here $\Psi(\omega)$ is absolutely integrable.

By Lemma IIh and Lemma IIg,

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} f(x+\omega)\,\Psi(\omega)\,d\omega - \sum_{k=-n^2}^{n^2-1} f\left(x+\frac{k}{n}\right) \int_{\frac{k}{n}}^{\frac{k+1}{n}} \Psi(\omega)\,d\omega \right| dx$$
$$= 0. \tag{2.53}$$

Combining this with (2.52) and Lemma IIi, we obtain Theorem II.

With only a slight modification of detail, we have proved:

THEOREM III. *Let $f_1(x)$ and $f_2(x)$ be two functions of L_1. Let*

$$g_1(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f_1(x)\,e^{iux}\,dx \tag{2.54}$$

and

$$g_2(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f_2(x)\,e^{iux}\,dx. \tag{2.55}$$

Let the set of points where $g_2(u) \neq 0$ consist only of inner points of the set where $g_1(u) \neq 0$. Then if $\epsilon > 0$ there is a polynomial

$$f_3(x) = \sum_1^N a_n f_1(x+\lambda_n) \tag{2.56}$$

such that

$$\int_{-\infty}^{\infty} |f_2(x) - f_3(x)|\,dx < \epsilon. \tag{2.57}$$

On the basis of Theorem III, we are in a position to prove the general:

THEOREM IV. *Let Σ be a class of functions of L_1. Then a necessary and sufficient condition for the class Σ_1, containing all functions $f(x+\lambda)$ whenever $f(x)$ belongs to Σ, to be closed L_1, is that there should be no real zero common to all the Fourier transforms of functions of Σ.*

The proof of the necessity of the condition of Theorem IV does not differ from the similar proof in the case of Theorem II. As to the sufficiency, it follows at once from Theorem III and the fact that the Fourier transform of a function of L_1 differs from 0 on an open set, that for every u, an ϵ may be assigned, such that

$$e^{iux}\frac{\sin^2 \epsilon x}{x^2} \tag{2.58}$$

is a function approximable L_1 by polynomials in functions of Σ_1. Let $(-U, U)$ be any interval of frequencies. By the Heine-Borel theorem, this may be overlaid with a finite number of overlapping intervals $U_n - \varepsilon_n$, $U_n + \varepsilon_n$. Let us form

$$\sum e^{iU_n x} \frac{\sin^2 \varepsilon_n x}{x^2} \tag{2.59}$$

for this set of intervals. This will be a function of L_1, approximable L_1 by polynomials in the function of Σ_1, whose Fourier transform may be shown to be

$$\sum \frac{\varepsilon_n}{4}\left(1-2\left|\frac{U_n+u}{\varepsilon_n}\right|\right)\left[\operatorname{sgn}\left(1-2\left|\frac{U_n+u}{\varepsilon_n}\right|\right)+1\right] \tag{2.60}$$

and to have no zeros over $(-U, U)$. Thus by Theorem III, every function of L_1 with a Fourier transform vanishing outside $(-U+n, U-n)$ will be approximable L_1 by polynomials in the functions of Σ_1. At this stage, the introduction of Lemma III serves to complete the proof of Theorem IV.

In Theorems II and IV, the necessary part is nearly trivial and all the difficulty resides in the proof of sufficiency. There are certain applications, however, where it is precisely the necessary part of the theorem that is significant. We may prove a function to have no zeros if it is the Fourier transform of a function of L_1 the set of whose translations is closed L_1. To establish this closure is indeed more than is needed: it is enough to produce for each u among the class of functions to which we may approximate L_1 by polynomials in the translations of our given functions, at least one function whose Fourier transform does not vanish for the argument u.

It may be shown if $F(x)$ is a non-negative function for which

$$\int_{-\infty}^{\infty} F(x)\,dx > A \tag{2.61}$$

and

$$\left[\int_B^{\infty} + \int_{-\infty}^{-B}\right] F(x)\,dx < \varepsilon A, \tag{2.62}$$

then

$$\left|\int_{-\infty}^{\infty} F(x)\,e^{iux}\,dx\right| > \left(\frac{1}{2} - \frac{3\varepsilon}{2}\right) A \qquad \left[|u| \leqq \frac{\pi}{3B}\right]. \tag{2.63}$$

This results from the fact that for $-B \leqq \lambda \leqq B$,

$$\Re(e^{iux}) = \cos ux \geqq \cos\frac{\pi}{3} = \frac{1}{2}.$$

Accordingly, if for every value of B and ε, we may find a function of this sort which may be approximated L_1 by polynomials in the translations of $f(x)$, the Fourier transform of $f(x)$ has no zeros. As in the

case of Theorem IV, the function $f(x)$ may be replaced by the class Σ, and the class of translations of $f(x)$ by Σ_1. Then the condition that the Fourier transform of $f(x)$ shall have no zeros is to be replaced by the condition that there shall be no zero common to all the Fourier transforms of functions of Σ.

This method is a conceivable one for the investigation of the zeros of functions with known Fourier transforms, and might be applied to the study of the zeros of the Riemann zeta function. So far it has yielded no results. Under certain conditions, the Taylor series may be introduced to reduce the question of the closure of the set of translations of $f(x)$ to that of the closure of the set of functions

$$f^{(n)}(x). \tag{2.65}$$

This matter has not yet, however, been subjected to an adequate investigation.

3. A sub-class of L_1. In connection with Stieltjes distributions of mass, we shall have to consider "kernels" of class L_1, fulfilling a certain more restrictive condition. This condition on a measurable function $f(x)$ is that the series

$$\sum_{k=-\infty}^{\infty} \overline{\lim_{kA+B \leq x \leq (k+1)A+B}} |f(x)| \tag{3.01}$$

shall converge. We may easily verify that this condition is in fact independent of A and B, and that it can only be fulfilled by functions of L_1. With it goes a certain definition of the "distance" between two functions $f(x)$ and $g(x)$, this "distance" being defined as

$$(f(x), g(x)) = \sum_{-\infty}^{\infty} \overline{\lim_{k \leq x \leq k+1}} |f(x)-g(x)|. \tag{3.02}$$

This notion of "distance" has a very important difference from

$$\int_{-\infty}^{\infty} |f(x)-g(x)|^2 dx \tag{3.03}$$

and

$$\int_{-\infty}^{\infty} |f(x)-g(x)| dx \tag{3.04}$$

in that it is not true that if (3.01) is finite, then

$$\lim_{\varepsilon \to 0} \sum_{k=-\infty}^{\infty} \overline{\lim_{k \leq x \leq k+1}} |f(x+\varepsilon)-f(x)| = 0. \tag{3.05}$$

Accordingly, if we are to prove theorems analogous to those in the last paragraph without a radical change in method of proof, we must introduce a restriction which will make (3.05) hold. Otherwise we shall be unable to establish the analogue of Lemma IIg.

A condition of the desired sort is that $f(x)$ shall be continuous. It will then be uniformly continuous over any finite interval, and we shall be able, first to choose A so large that

$$\sum_{k=0}^{\infty} \left\{ \max_{A+k\leqq x\leqq A+k+1} |f(x)| + \max_{-A-k-1\leqq x\leqq -A-k} |f(x)| \right\} < \varepsilon \tag{3.06}$$

and then to choose η so small that for $-A-1 \leqq x \leqq A+1$, $0<\eta_1<\eta$,

$$\max |f(x+\eta_1)-f(x)| < \frac{\varepsilon}{2A+2}. \tag{3.07}$$

It will follow that for $0<\eta_1<\eta$,

$$\sum_{k=-\infty}^{\infty} \max_{k\leqq x\leqq k+1} |f(x+\varepsilon)-f(x)| < 3\varepsilon. \tag{3.08}$$

We shall call the class of all continuous functions for which (3.01) is finite, the class M_1, and we shall say that a class Σ of functions of M_1 is *closed* M_1, if whenever $F(x)$ belongs to M_1 and $\varepsilon>0$, there is a polynomial

$$F_1(x) = \sum_{h=1}^{n} A_h f_h(x) \tag{3.09}$$

in functions of the class Σ such that

$$\sum_{k=-\infty}^{\infty} \max_{k\leqq x\leqq k+1} |F(x)-F_1(x)| < \varepsilon. \tag{3.10}$$

The following theorem is valid:

THEOREM V. *Let $f(x)$ belong to L_1. Then a necessary and sufficient condition that whenever $f_1(x)$ is a function of M_1 and $\varepsilon>0$, there shall exist a function $\alpha(x)$ of M_1 such that*

$$\sum_{k=-\infty}^{\infty} \max_{k\leqq x\leqq k+1} \left| f_1(x) - \int_{-\infty}^{\infty} f(y)\alpha(x-y)\,dy \right| < \varepsilon \tag{3.11}$$

is that for no real u

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{iux}\,dx = 0. \tag{3.12}$$

The proof differs in no essential respect from that of Theorem II. Let us note that if $f_2(x)$ belongs to L_1 and

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f_2(x)\,e^{iux}\,dx = 0, \qquad [|u|>A] \tag{3.13}$$

then $f_2(x)$ belongs to M_1. To see this, let us put

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f_2(x)\,e^{-iux}\,dx = g_2(u); \tag{3.14}$$

$g_2(u)$ is bounded and differs from 0 only over a bounded range, and hence belongs to L_2. Thus

$$\underset{B\to\infty}{\text{l.i.m.}}\ \frac{1}{\sqrt{2\pi}}\int_{-B}^{B} g_2(u)\,e^{iux}\,du = f_3(x) \tag{3.15}$$

exists as a function of L_2. Further

$$g_2(u) = \underset{B\to\infty}{\text{l.i.m.}}\ \frac{1}{\sqrt{2\pi}}\int_{-B}^{B} f_3(x)\,e^{-iux}\,dx \tag{3.16}$$

exists and belongs to L_2. If we put

$$F(x) = \int_{-\infty}^{\infty} G(u)\,e^{iux}\,du \tag{3.17}$$

where G belongs both to L_2 and L_1, it follows readily from (3.14) and (3.16) that

$$\int_{-\infty}^{\infty} F(x)\,[f_2(x)-f_3(x)]\,dx = 0. \tag{3.18}$$

We can choose a set of F's vanishing outside a given finite interval and closed L_2 and hence L_1 over this interval. Then over this interval, which is arbitrary, f_2 and f_3 can differ at most on a null set. Thus f_2 belongs to L_2.

Now

$$g_2(u) = g_2(u)\,W(u) \tag{3.19}$$

where

$$W(u) = \begin{cases} 0 & ; \quad [|u|>2A] \\ 2-\dfrac{|u|}{A} & ; \quad [A<|u|<2A] \\ 1 & . \quad [|u|<A] \end{cases} \tag{3.20}$$

Thus by the Parseval theorem,

$$\begin{aligned} f_2(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} f_2(y)\,dy\int_{-\infty}^{\infty} W(u)\,e^{iu(x-y)}\,du \\ &= \frac{1}{\pi A}\int_{-\infty}^{\infty} f_2(y)\,\frac{\cos A(x-y)-\cos 2A(x-y)}{(x-y)^2}\,dy. \end{aligned} \tag{3.21}$$

From this it follows that

$$\sum_{k=-\infty}^{\infty} \max_{k\leqq x\leqq k+1} |f_2(x)| \leqq \frac{1}{\pi A}\int_{-\infty}^{\infty} |f_2(y)|\,dy \sum_{-\infty}^{\infty} O\left(\frac{1}{k-m}\right)^2 \tag{3.22}$$
$$\leqq O(1)\int_{-\infty}^{\infty} |f_2(y)|\,dy$$

and that $f_2(x)$, which obviously can be modified so as to be continuous, belongs to M_1.

It only remains to prove the analogue of Lemma IIi, and to show that every function of class M_1 may be approximated with any degree of accuracy in the M_1 sense by functions with Fourier transforms vanishing for large arguments. This follows exactly the lines suggested by IIi. We wish to show, that is, that if $f(x)$ belongs to M_1,

$$\lim_{N\to\infty}\left(f(x), \frac{1}{\pi N}\int_{-\infty}^{\infty} f(x+y)\frac{\sin^2 Ny}{y^2}\,dy\right) \equiv 0. \tag{3.221}$$

As before, we put

$$\Phi_1(y) = \begin{cases} \dfrac{\sin^2 Ny}{y^2}(1-|y|N^{1/2}); & [|y| \leqq N^{-1/2}] \\ 0 \qquad\qquad . & [y > N^{-1/2}] \end{cases} \tag{3.222}$$

$$\Phi_2(y) = \frac{\sin^2 Ny}{y^2} - \Phi_1(y). \tag{3.223}$$

The proof that

$$\lim_{N\to\infty}\left(f(x), \frac{1}{\pi N}\int_{-\infty}^{\infty} f(x+y)\,\Phi_1(y)\,dx\right) = 0 \tag{3.224}$$

follows exactly the lines of that of (2.41), for the analogue of Lemma IIg has already been shown to be true, while that of Lemma IIh is proved by the same arguments of absolute convergence as the lemma itself. Again,

$$\lim_{N\to\infty} \sum_{k=-\infty}^{\infty} \max_{k\leqq x\leqq k+1} \left|\frac{1}{\pi N}\int_{-\infty}^{\infty} f(x+y)\,\Phi_2(y)\,dy\right|$$
$$\leqq \sum_{k=-\infty}^{\infty} \max_{k\leqq x\leqq k+1} |f(x+y)| \lim_{N\to\infty}\frac{1}{\pi N}\int_{-\infty}^{\infty} \Phi_2(y)\,dy \tag{3.225}$$
$$\leqq 2\sum_{k=-\infty}^{\infty} \max_{k\leqq x\leqq k+1} |f(x)| \lim_{N\to\infty}\frac{1}{\pi N}\int_{-\infty}^{\infty} \Phi_2(y)\,dy,$$

and this, by (2.421), is zero. This completes the proof of (3.221).

An extension of Theorem V in the direction of Theorem IV is the following:

THEOREM VI. *Let Σ be a class of functions of L_1. Let Σ_1 be the class of all functions of the form*

$$\sum_1^N \int_{-\infty}^{\infty} f_n(y)\, \alpha_n(x+y)\, dy \tag{3.23}$$

where $f_1, \cdots, f_N$ belong to $\sum$ and $\alpha_1, \cdots, \alpha_N$ to M_1. Then Σ_1 is closed M_1 when and only when there is no real value of u which is a zero common to all the functions

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f_n(x)\, e^{iux}\, dx \tag{3.24}$$

corresponding to functions $f_n(x)$ of class Σ.

As an easy corollary of Theorem V, using (3.10), we have:

THEOREM VII. *If $f(x)$ is a function of M_1, a necessary and sufficient condition for the set of its translations to be closed M_1 is that for no u should we have*

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x)\, e^{iux}\, dx = 0. \tag{3.25}$$

CHAPTER II.

ASYMPTOTIC PROPERTIES OF AVERAGES[26].

4. Averages of bounded functions. Our fundamental theorem is:

THEOREM VIII. *Let $f(x)$ be a bounded measurable function, defined over $(-\infty, \infty)$. Let $K_1(x)$ be a function in L_1, and let*

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} K_1(x)\, e^{iux}\, dx \neq 0 \tag{4.01}$$

for every real u. Let

$$\lim_{x\to\infty} \int_{-\infty}^{\infty} f(\xi)\, K_1(\xi - x)\, d\xi = A \int_{-\infty}^{\infty} K_1(\xi)\, d\xi. \tag{4.02}$$

Then if $K_2(x)$ is any function in L_1,

$$\lim_{x\to\infty} \int_{-\infty}^{\infty} f(\xi)\, K_2(\xi - x)\, d\xi = A \int_{-\infty}^{\infty} K_2(\xi)\, d\xi. \tag{4.03}$$

Conversely, let $K_1(\xi)$ be a function of L_1, and let $\int_{-\infty}^{\infty} K_1(\xi)\, d\xi \neq 0$. Let (4.02) *imply* (4.03) *whenever $K_2(x)$ belongs to L_1 and $f(x)$ is bounded. Then* (4.01) *holds.*

[26] The emphasis here placed on averages is in the same order of ideas as was first introduced into the theory by Schmidt (1) (2).

As to the first part of Theorem VIII, it is clear that it is valid whenever $K_2(x)$ is of the form $\sum_{k=1}^{n} A_k K_1(x+\lambda_k)$. It is also clear that if

$$|f(x)| \leqq B; \qquad \int_{-\infty}^{\infty} |K_2(x) - K_3(x)|\, dx \leqq \epsilon, \tag{4.04}$$

then

$$\left| \int_{-\infty}^{\infty} f(\xi)\, K_2(\xi - x)\, d\xi - \int_{-\infty}^{\infty} f(\xi)\, K_3(\xi - x)\, d\xi \right| \leqq B\epsilon.$$

An application of Theorem II completes the proof.

As to the second part of the proof, it is merely necessary to suppose that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} K_1(x)\, e^{iu_1 x}\, dx = 0 \tag{4.05}$$

and to take

$$f(x) = e^{iu_1 x} \tag{4.06}$$

to obtain a contradiction.

Theorem VIII has an extension in the sense of Theorem III, in which $K_1(x)$ is replaced by a class of functions Σ_1 (4.02) holds for every function of the class, and there is no u for which for *every* function $K_1(x)$ of Σ_1

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} K_1(x)\, e^{iux}\, dx = 0. \tag{4.07}$$

The converse of this extended theorem is also valid.

5. **Averages of bounded Stieltjes distributions.** Here our theorem is:

THEOREM IX. *Let $f(x)$ be a function of limited total variation over every finite range, and let*

$$\int_{y}^{y+1} |df(x)| \tag{5.01}$$

be bounded in y. Let $K_1(x)$ be a continuous function of L_1, and let

$$\sum_{k=-\infty}^{\infty} \overline{\lim_{k \leqq x \leqq k+1}} |K_1(x)| \tag{5.02}$$

converge. Let

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} K_1(x) e^{iux}\, dx \neq 0 \qquad [-\infty < u < \infty] \tag{4.01}$$

and let

$$\lim_{x \to \infty} \int_{-\infty}^{\infty} K_1(\xi - x)\, df(\xi) = A \int_{-\infty}^{\infty} K_1(\xi)\, d\xi. \tag{5.03}$$

Then if $K_2(x)$ is any function of M_1,

$$\lim_{x \to \infty} \int_{-\infty}^{\infty} K_2(\xi - x)\, df(\xi) = A \int_{-\infty}^{\infty} K_2(\xi)\, d\xi. \tag{5.04}$$

Conversely, let (5.02) *converge. Let* $\int_{-\infty}^{\infty} K_1(x)\,dx \neq 0$. *Let* (5.03) *imply* (5.04) *for every function of* K_2 *of* M_1 *and every* $f(x)$ *for which* (5.01) *is bounded. Then* (4.01) *holds.*

To prove this theorem we need the following elementary:

LEMMA IXa. *Let*

$$\int_{y-0}^{y+1+0} |df(x)| \leqq B \qquad [-\infty < y < \infty] \tag{5.05}$$

and let

$$\sum_{k=-\infty}^{\infty} \overline{\lim_{k \leqq x \leqq k+1}} |K_2(x) - K_3(x)| \leqq \varepsilon. \tag{5.06}$$

Then

$$\left| \int_{-\infty}^{\infty} K_2(\xi - x)\,df(x) - \int_{-\infty}^{\infty} K_3(\xi - x)\,df(x) \right| \leqq B\varepsilon. \tag{5.061}$$

With this lemma at our disposal, the proof of Theorem IX does not differ in any important respect from that of Theorem VIII. Of course, as in Theorem VI, we must replace the polynomials in translations of K_1 that figure in Theorem VIII by absolutely convergent integrals in these translations. The proof of the lemma itself is immediate.

The modification of Theorem IX with a hypothesis involving a whole class of kernels will be used later, so we shall formulate it as a separate theorem:

THEOREM X. *Let* $f(x)$ *be a function of limited total variation over every finite range, and let*

$$\int_{y}^{y+1} |df(x)| \tag{5.01}$$

be bounded in y. *Let* Σ *be a class of continuous functions of* L_1, *each one of which, for example* $K_1(x)$, *has the properties that*

$$\sum_{k=-\infty}^{\infty} \overline{\lim_{k \leqq x \leqq k+1}} |K_1(x)| \tag{5.02}$$

converges, and that

$$\lim_{x \to \infty} \int_{-\infty}^{\infty} K_1(\xi - x)\,df(\xi) = A \int_{-\infty}^{\infty} K_1(\xi)\,d\xi. \tag{5.03}$$

Let there be no u *which is a real zero for all the functions*

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} K_1(x) e^{iux}\,dx \tag{5.07}$$

for which $K_1(x)$ *belongs to* Σ. *Then if* $K_2(x)$ *is any function belonging to* M_1,

$$\lim_{x\to\infty}\int_{-\infty}^{\infty} K_2(\xi-x)\,df(\xi) = A\int_{-\infty}^{\infty} K_2(\xi)\,d\xi. \tag{5.04}$$

Conversely, if the class $\sum$ *contains at least one function with non zero integral and has the property that each member of it is a member of* L_1 *and satisfies the condition that* (5.02) *shall converge, and if, whenever* $f(x)$ *satisfies the condition that* (5.01) *be bounded, and* (5.03) *for every member of* Σ, *then* (5.04) *holds for every function* $K_2(x)$ *belonging to* M_1, *then there is no* u *such that* (4.07) *holds for every function* $K_1(x)$ *belonging to* Σ.

6. Averages of unilaterally bounded distributions and functions. Let $K_1(x)$ not be equivalent to 0, and let it be a continuous function (or a function continuous except for a finite number of finite jumps) not identically zero and such that

$$K_1(x) \geqq 0, \quad \sum_{k=-\infty}^{\infty} \max_{k\leqq x\leqq k+1} K_1(x) < p \qquad (-\infty < x < \infty). \tag{6.01}$$

We shall say that the mass-distribution determined by $f(x)$ is *bounded below* when

$$\int_y^{y+1} |df(x)| - f(y+1) + f(y) \leqq N \qquad (-\infty < x < \infty) \tag{6.02}$$

and that it is *bounded above* when $-f(x)$ determines a mass-distribution bounded below. We shall prove:

LEMMA XIa. *Let the distribution corresponding to* $f(x)$ *be bounded below* (*or above*). *Let* (6.01) *hold, and let*

$$\bar{\int}_{-\infty}^{\infty} K_1(\xi-x)\,df(\xi) \leqq M; \quad (\geqq M); \qquad (-\infty < x < \infty). \tag{6.03}$$

Then there is a Q *such that*

$$\int_y^{y+1} |df(x)| \leqq Q \qquad (-\infty < y < \infty). \tag{6.04}$$

To prove this, let us notice that

$$\begin{aligned}\bar{\int}_{-\infty}^{\infty} K_1(\xi-x)\,d\int_{\xi_0}^{\xi} |df(y)| &\leqq \bar{\int}_{-\infty}^{\infty} K_1(\xi-x)\,d\left[\int_{\xi_0}^{\xi} |df(y)| - f(\xi)\right] \\ &+ \bar{\int}_{-\infty}^{\infty} K_1(\xi-x)\,df(\xi) \leqq N\sum_{k=-\infty}^{\infty} \max_{k\leqq x\leqq k+1} K(\xi) + M.\end{aligned} \tag{6.05}$$

Since K is "stückweise stetig", B, a and b exist such that

$$K_1(u) > B > 0 \qquad (a \leqq u \leqq b). \tag{6.06}$$

Hence

$$M + N\sum_{k=-\infty}^{\infty} \max_{k\leqq x\leqq k+1} K_1(\xi) \geqq B\int_{a+x}^{b+x} d\int_{\xi_0}^{\xi} |df(y)|, \tag{6.07}$$

or in other terms

$$\frac{1}{B}\left\{\left[\frac{1}{b-a}\right]+1\right\}\left\{M+N\sum_{k=-\infty}^{\infty}\max_{k\leq x\leq k+1}K_1(\xi)\right\} \geqq \int_y^{y+1}|df(x)| \qquad (-\infty<y<\infty). \tag{6.08}$$

If we combine this lemma with Theorem X, we get:

THEOREM XI. *Let $f(x)$ be a function of limited total variation over every finite interval, for which*

$$\int_y^{y+1}|df(x)|-f(y+1)+f(y)\leqq N \qquad (-\infty<y<\infty). \tag{6.02}$$

Let Σ be a class of continuous functions K_1, for each of which

$$\sum_{k=-\infty}^{k=\infty}\max_{k\leq x\leq k+1}|K_1(x)| \tag{5.02}$$

converges, and let

$$\lim_{x\to\infty}\int_{-\infty}^{\infty}K_1(\xi-x)\,df(\xi) = A\int_{-\infty}^{\infty}K_1(\xi)\,d\xi \tag{5.03}$$

for every K_1 belonging to Σ. Let there be no real u for which every

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}K_1(x)\,e^{iux}\,dx \tag{5.07}$$

vanishes, for which K_1 belongs to Σ. Let $Q(x)$ be a continuous function belonging to Σ, for which

$$Q(x)\geqq 0, \quad \left|\int_{-\infty}^{\infty}Q(\xi-x)\,df(x)\right|\leqq T \qquad (-\infty<x<\infty). \tag{6.09}$$

Then if $K_2(-x)$ is any function belonging to M_1,

$$\lim_{x\to\infty}\int_{-\infty}^{\infty}K_2(\xi-x)\,df(\xi) = A\int_{-\infty}^{\infty}K_2(\xi)\,d\xi. \tag{5.04}$$

We shall have frequent occasion to use Theorems X and XI in a form in which the infinite range of x is mapped on a semi-infinite range by an exponential transformation. This may happen in two ways: either 0 or ∞ for the new argument may correspond to $+\infty$ for x. In the first case, let us write:

$$\xi=-\log\lambda,\ f(\xi)=\int\lambda^{-1}\,d\varphi(\lambda),\ K_{1,2}(\xi)=\lambda N_{1,2}(\lambda),\ Q(\xi)=\lambda M(\lambda) \tag{6.10}$$

and in the second,

$$\xi=\log\lambda,\ f(\xi)=\int\lambda^{-1}\,d\varphi(\lambda),\ K_{1,2}(\xi)=\lambda N_{1,2}(\lambda),\ Q(\xi)=\lambda M(\lambda). \tag{6.11}$$

Then Theorem XI yields:

THEOREM XI′. *Let $\varphi(\lambda)$ be a function of limited total variation over every interval $(\varepsilon, 1/\varepsilon)$ where $0<\varepsilon<1$. Let $\varphi(0) = 0$ and let*

$$\int_u^{2u} \lambda^{-1} \, |\, d\varphi(\lambda)\,| - \int_u^{2u} \lambda^{-1} \, d\varphi(\lambda) \leqq N \qquad (0<u<\infty). \tag{6.12}$$

Let Σ be a class of continuous functions $N_1(\lambda)$ for each of which

$$\sum_{k=-\infty}^{\infty} \max_{2^k \leqq \lambda \leqq 2^{k+1}} \lambda \, |\, N_1(\lambda)\,| \tag{6.13}$$

converges, and let

$$\lim_{\lambda \to 0\,(\infty)} \frac{1}{\lambda} \int_0^{\infty} N_1\left(\frac{\mu}{\lambda}\right) d\varphi(\mu) = A \int_0^{\infty} N_1(\mu)\, d\mu \tag{6.14}$$

for every $N_1(\lambda)$ belonging to Σ. Let there be no real u for which every

$$\int_0^{\infty} N_1(\lambda)\, \lambda^{iu}\, d\lambda \tag{6.15}$$

vanishes, for which N_1 belongs to Σ. Let $M(\lambda)$ be a function belonging to Σ, for which

$$M(\lambda) \geqq 0, \quad \left| \frac{1}{\lambda} \int_0^{\infty} M\left(\frac{\mu}{\lambda}\right) d\varphi(\mu) \right| \leqq \text{const. } \textit{for } 0 \leqq \lambda < \infty. \tag{6.16}$$

Then if $N_2(\lambda)$ is any continuous function for which

$$\sum_{k=-\infty}^{\infty} \max_{2^k \leqq \lambda \leqq 2^{k+1}} \lambda \, |\, N_2(\lambda)\,| \tag{6.17}$$

converges,

$$\lim_{\lambda \to 0\,(\infty)} \frac{1}{\lambda} \int_0^{\infty} N_2\left(\frac{\mu}{\lambda}\right) d\varphi(\mu) = A \int_0^{\infty} N_2(\mu)\, d\mu. \tag{6.18}$$

In the case where

$$\varphi(\lambda) = \int_{\lambda_0}^{\lambda} f(x)\, dx, \tag{6.19}$$

(6.12) will be satisfied if

$$f(\lambda) > -\frac{N}{2 \log 2}. \tag{6.20}$$

If the other hypotheses of Theorem XI′ are satisfied, (6.18) will become

$$\lim_{\lambda \to 0(\infty)} \frac{1}{\lambda} \int_0^{\infty} N_2\left(\frac{\mu}{\lambda}\right) f(\mu)\, d\mu = A \int_0^{\infty} N_2(\mu)\, d\mu. \tag{6.21}$$

As a particular admissible N_2, we may take

$$N_2(\mu) = \begin{cases} 1 & ; \quad [0 \leqq \mu < 1] \\ 1 - \dfrac{\mu - 1}{\varepsilon} & ; \quad [1 \leqq \mu < 1+\varepsilon] \\ 0 & . \quad [1+\varepsilon \leqq \mu] \end{cases} \tag{6.22}$$

Thus (6.21) will assume the form

$$\lim_{\lambda \to 0(\infty)} \frac{1}{\lambda} \left[\int_0^{\lambda} f(\mu)\, d\mu + \int_{\lambda}^{\lambda(1+\varepsilon)} \left(1 - \frac{\mu - \lambda}{\lambda \varepsilon}\right) f(\mu)\, d\mu \right] = A\left(1 + \frac{\varepsilon}{2}\right). \tag{6.23}$$

Now, by (6.20)

$$\frac{1}{\lambda} \int_{\lambda}^{\lambda(1+\varepsilon)} \left(1 - \frac{\mu - \lambda}{\lambda \varepsilon}\right) f(\mu)\, d\mu \geqq \frac{1}{\lambda} \int_{\lambda}^{\lambda(1+\varepsilon)} \left(1 - \frac{\mu - \lambda}{\lambda \varepsilon}\right) \left(\frac{-N}{2 \log 2}\right) d\mu \tag{6.24}$$

$$= \frac{-N\varepsilon}{4 \log 2}.$$

Hence by (6.23)

$$\overline{\lim_{\lambda \to 0(\infty)}} \frac{1}{\lambda} \int_0^{\lambda} f(\mu)\, d\mu \leqq A\left(1 + \frac{\varepsilon}{2}\right) + \frac{N\varepsilon}{4 \log 2}. \tag{6.25}$$

Since ε is arbitrarily small,

$$\overline{\lim_{\lambda \to 0(\infty)}} \frac{1}{\lambda} \int_0^{\lambda} f(\mu)\, d\mu \leqq A. \tag{6.25}$$

Again, we may write (6.21) in the form

$$\lim_{\lambda \to 0(\infty)} \frac{1}{\lambda(1+\varepsilon)} \left[\int_0^{\lambda(1+\varepsilon)} f(\mu)\, d\mu - \int_{\lambda}^{\lambda(1+\varepsilon)} \frac{\mu - \lambda}{\lambda \varepsilon} f(\mu)\, d\mu \right] = \frac{A(1+\varepsilon/2)}{1+\varepsilon} \tag{6.26}$$

from which we may conclude as above that

$$\underline{\lim_{\lambda \to 0(\infty)}} \frac{1}{\lambda} \int_0^{\lambda} f(\mu)\, d\mu \geqq A. \tag{6.27}$$

Combining (6.25) and (6.27), we see that

$$\lim_{\lambda \to 0(\infty)} \frac{1}{\lambda} \int_0^{\lambda} f(\mu)\, d\mu = A. \tag{6.28}$$

We thus get

THEOREM XI″. *Let $f(x)$ be a function bounded over every interval $(\varepsilon, 1/\varepsilon)$, where $0 < \varepsilon < 1$. Let*

$$f(\lambda) > -K \text{ (or } f(\lambda) < K) \tag{6.29}$$

for every argument. Let Σ be a class of continuous functions $N_1(\lambda)$ for each of which (6.13) *converges, and let*

$$\lim_{\lambda\to 0(\infty)} \frac{1}{\lambda}\int_0^\infty N_1\left(\frac{\mu}{\lambda}\right) f(\mu)\,d\mu = A\int_0^\infty N_1(\mu)\,d\mu \tag{6.30}$$

for every $N_1(\lambda)$ *belonging to* Σ. *Let there be no real* u *for which every expression* (6.15) *vanishes with* N_1 *belonging to* Σ. *Let* $M(\lambda)$ *belong to* Σ, *and let*

$$M(\lambda) \geqq 0, \frac{1}{\lambda}\int_0^\infty M\left(\frac{\mu}{\lambda}\right) f(\mu)\,d\mu \leqq \text{const. } \textit{for } 0 \leqq \lambda < \infty\,. \tag{6.31}$$

Then (6.28) *is true.*

Another case where a stricter conclusion than (6.18) may be drawn is where $\varphi(\lambda)$ is monotone. In this case, (6.12) is automatically satisfied. If we take $N_2(\mu)$ as in (6.22), we shall be able to write (6.18) in the form

$$\lim_{\lambda\to 0(\infty)} \frac{1}{\lambda}\left[\varphi(\lambda) + \int_\lambda^{\lambda(1+\varepsilon)} \left(1 - \frac{\mu-\lambda}{\lambda\varepsilon}\right) d\varphi(\mu)\right] = A\left(1+\frac{\varepsilon}{2}\right) \tag{6.32}$$

and hence

$$\overline{\lim_{\lambda\to 0(\infty)}} \frac{\varphi(\lambda)}{\lambda} \leqq A. \tag{6.33}$$

From the analogue of (6.26) it follows that

$$\underline{\lim_{\lambda\to 0(\infty)}} \frac{\varphi(\lambda)}{\lambda} \geqq A, \tag{6.34}$$

and hence that

$$\varphi(\lambda) \sim \lambda A. \tag{6.35}$$

We thus obtain

THEOREM XI‴. *If in the hypothesis of Theorem* XI′, (6.12) *is replaced by the condition that* $\varphi(\lambda)$ *is monotone. Then* (6.35) *follows.*

It is even possible to weaken the hypothesis here given, and to replace (6.12) by the condition that

$$\lim_{\varepsilon\to+0} \overline{\lim_{-\infty<\mu<\infty}} \left[\int_\mu^{\mu(1+\varepsilon)} \lambda^{-1}\,|d\varphi(\lambda)| - \int_\mu^{\mu(1+\varepsilon)} \lambda^{-1}\,d\varphi(\lambda)\right] = 0 \tag{6.36}$$

to establish (6.25).

CHAPTER III.

TAUBERIAN THEOREMS AND THE CONVERGENCE OF SERIES AND INTEGRALS.

7. The Hardy-Littlewood condition. We now enter upon the realm of ideas which has longest been associated with Tauberian theorems.[27] The class Σ of the last paragraph now consists of all functions of the form

[27] Hardy (2); Hardy and Littlewood (4), (10), (13) etc.; Littlewood (1); Landau (9), (4).

(7.01) $$M(x) - M(x+\lambda)$$

for which we have

(7.02) $M(-\infty) = 1, \quad M(\infty) = 0, \quad \int_{-\infty}^{\infty} M(\xi - x)\, df(\xi)$ bounded.

Condition (5.02) now asserts the convergence of

(7.03) $$\sum_{k=-\infty}^{\infty} \max_{k \leqq x \leqq k+1} |M(x) - M(x+\lambda)|$$

which will be automatically fulfilled if $M(x)$ is monotone and satisfies (7.02). We shall also suppose $M(x)$ continuous. Condition (5.07) becomes the condition that

(7.04) $$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} [M(x) - M(x+\lambda)]\, e^{iux}\, dx$$

shall not vanish for any real u for every λ. Let us suppose M to vanish exponentially at $+\infty$, and let

(7.05) $$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} M(x)\, e^{xz}\, dx = \varphi(z).$$

$\varphi(z)$ will clearly be analytic over some vertical strip to the right of the origin. We shall obviously have

(7.06) $$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} [M(x) - M(x+\lambda)]\, e^{xz}\, dz = \varphi(z)\,(1 - e^{-\lambda z})$$

over this strip, and by analytic continuation, which we assume to be possible, the non-vanishing of (7.04) becomes the non-vanishing of $\varphi(z)$ over the imaginary axis.

If

(7.07) $$\lim_{x \to \infty} \int_{-\infty}^{\infty} M(\xi - x)\, df(\xi) = B,$$

(5.03) is satisfied for $A = 0$. As is obviously permissible, let us put

(7.08) $$K_2(\xi) = \begin{cases} M(\xi) - 1 & ; \quad (-\infty < \xi < 0) \\ M(\xi) - 1 + \dfrac{\xi}{\varepsilon} & ; \quad (0 \leqq \xi < \varepsilon) \\ M(\xi) & . \quad (\varepsilon \leqq \xi < \infty) \end{cases}$$

Then (5.04) assumes the form

(7.09) $$\lim_{x \to \infty} \int_{-\infty}^{\infty} K_2(\xi - x)\, df(\xi) = 0,$$

which becomes

(7.10) $$\lim_{x \to \infty} \left[\int_{-\infty}^{x} df(\xi) + \int_{x}^{x+\varepsilon} \left(1 - \frac{\xi - x}{\varepsilon}\right) df(\xi) \right] = B.$$

If we put

$$f(-\infty) = 0 \tag{7.11}$$

(5.04) assumes the form

$$\lim_{x\to\infty} \frac{1}{\varepsilon}\int_x^{x+\varepsilon} f(\xi)\,d\xi = B. \tag{7.12}$$

Restating Theorem XI in this new form, we get

THEOREM XII. *Let $f(x)$ be a function of limited total variation over every range $(-\infty, A)$, let $f(-\infty) = 0$, and let*

$$\int_y^{y+1} |\,df(x)\,| - f(y+1) + f(y) \leqq N \qquad (-\infty < y < \infty). \tag{6.02}$$

Let $M(x)$ be a monotonely decreasing continuous function, such that

$$M(-\infty) = 1, \quad M(x) = O(e^{-ux}) \text{ at } \infty \qquad (u > 0). \tag{7.13}$$

Let

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} M(x)\,e^{xz}\,dx = \varphi(z) \tag{7.05}$$

over a strip to the right of the origin, and let $\varphi(z)$, continued analytically on to the imaginary axis, have no zeros there. (We assume the possibility of this continuation.) *Let*

$$\int_{-\infty}^{\infty} M(\xi - x)\,df(\xi) \tag{7.14}$$

be bounded, and let

$$\lim_{x\to\infty}\int_{-\infty}^{\infty} M(\xi - x)\,df(\xi) = B. \tag{7.07}$$

Then

$$\lim_{x\to\infty} \frac{1}{\varepsilon}\int_x^{x+\varepsilon} f(\xi)\,d\xi = B. \tag{7.12}$$

Let us put

$$M(\log x) = N(x); \quad f(\log x) = F(x).$$

Theorem XII then becomes:

THEOREM XIII. *Let $F(x)$ be a function vanishing at the origin and of limited total variation over any finite range (including the origin), and let for some $A > 1$,*

$$\int_y^{Ay} |\,dF(x)\,| - F(Ay) + F(y) \leqq N \qquad (0 < y < \infty). \tag{7.16}$$

Let $N(x)$ be a monotonely decreasing continuous function, such that

$$N(0) = 1, \qquad N(x) = O(x^{-\lambda}) \text{ at } \infty \qquad [\lambda > 0]. \tag{7.17}$$

Let

$$\frac{1}{\sqrt{2\pi}}\int_0^{\infty} N(x)\,x^{z-1}\,dx = \varphi(z) \tag{7.18}$$

over a strip to the right of the origin, and let $\varphi(z)$, continued analytically on to the imaginary axis, as we assume to be possible for all points but the origin, have no zeros there. Let

$$\int_0^\infty N\left(\frac{\xi}{x}\right) dF(\xi) \tag{7.19}$$

be bounded, and let

$$\lim_{x\to\infty} \int_0^\infty N\left(\frac{\xi}{x}\right) dF(\xi) = B. \tag{7.20}$$

Then if $\Lambda_1 > 1$,

$$\lim_{x\to\infty} \frac{1}{\log \Lambda_1} \int_x^{x\Lambda_1} F(\xi) \frac{d\xi}{\xi} = B. \tag{7.21}$$

A case of particular interest and importance is where

$$F(x) = \sum_{n=0}^{[x]} a_n. \tag{7.22}$$

Here

$$\begin{aligned} \int_y^{\Lambda y} |dF(x)| - F(\Lambda y) + F(y) &= \sum_{n=[y]+1}^{[\Lambda y]} (-a_n + |a_n|) \\ &\leqq \begin{cases} \max 2n a_n; \ (n \leqq [\Lambda y]), \\ 0. \end{cases} \end{aligned} \tag{7.23}$$

if Λ is sufficiently near to 1. (7.23) is to be interpreted as meaning that the expression on the left hand side is less than or equal to the greater of the two expressions in brackets.

We thus obtain the

COROLLARY. *If $N(x)$ is subject to the conditions of Theorem* XIII, *if*

$$\int_0^\infty N\left(\frac{\xi}{x}\right) dF(\xi) = \sum_{n=0}^\infty a_n N\left(\frac{n}{x}\right) \tag{7.24}$$

is bounded, if

$$n a_n < K \quad (\text{or } n a_n > -K) \tag{7.25}$$

for all n, and if

$$\lim_{x\to\infty} \sum_{n=0}^\infty a_n N\left(\frac{n}{x}\right) = B \tag{7.26}$$

then

$$\sum_0^\infty a_n = B. \tag{7.27}$$

Here instead of (7.27), what we directly prove is

$$\lim_{x\to\infty} \frac{1}{\log \Lambda_1} \int_x^{x\Lambda_1} \sum_{n=0}^{[\xi]} a_n \frac{d\xi}{\xi} = B. \tag{7.28}$$

(7.27) may however be written in the form

$$\lim_{x\to\infty}\frac{1}{\log \Lambda_1}\int_x^{x\Lambda_1}\sum_{n=0}^{[x]} a_n\frac{d\xi}{\xi} = B. \tag{6.281}$$

Our conclusion (7.27) will then follow if we can show that we can so choose Λ_1, that

$$\begin{aligned}\overline{\lim_{x\to\infty}}\ \frac{1}{\log \Lambda_1}\int_x^{x\Lambda_1}\sum_{n=[x]+1}^{[\xi]} a_n\frac{d\xi}{\xi} &< \varepsilon,\\ \overline{\lim_{x\to\infty}}\ \frac{1}{\log \Lambda_1}\int_x^{x\Lambda_1}\sum_{n=[\xi]+1}^{[x\Lambda_1]} a_n\frac{d\xi}{\xi} &< \varepsilon.\end{aligned} \tag{7.29}$$

However,

$$\frac{1}{\log \Lambda_1}\int_x^{x\Lambda_1}\sum_{n=[x]+1}^{[\xi]} a_n\frac{d\xi}{\xi} \leqq \frac{1}{\log \Lambda_1}\int_x^{x\Lambda_1}\frac{K}{x}(x\Lambda_1-x)\frac{d\xi}{\xi} = K(\Lambda_1-1). \tag{7.30}$$

Again,

$$\frac{1}{\log \Lambda_1}\int_x^{x\Lambda_1}\sum_{n=[\xi]+1}^{[\Lambda_1 x]} a_n\frac{d\xi}{\xi} \leq K(\Lambda_1-1). \tag{7.31}$$

Since we may take Λ_1 as near to 1 as we wish, (7.29) is established, and (7.27) follows at once. This type of argument is to be found in the work of Szász.

8. **The Schmidt condition.**[28] This last corollary covers the work of Hardy and Littlewood on Tauberian theorems. An extension of their conditions is due to Robert Schmidt. To arrive at theorems of his type, let M be subject to the conditions stated in the hypothesis to Theorem XII; let us put

$$g(x) = f(x+\alpha)-f(x) \tag{8.01}$$

and let us assume that

$$-f(x)+|f(x)| < A(x+|x|)+c \qquad (A, c>0). \tag{8.02}$$

As the integrals in question converge absolutely by (8.02) and (7.13), whenever they exist we may invert our order of integration and write:

$$\begin{aligned}\int_y^{y+\alpha} dx\int_{-\infty}^{\infty} M(\xi-x)\,df(\xi) &= \int_y^{y+\alpha} dx\int_{-\infty}^{\infty} M(\xi)\,df(\xi+x)\\ &= \int_{-\infty}^{\infty} M(\xi)\big(f(\xi+y+\alpha)-f(\xi+y)\big)\,d\xi\\ &= \int_{-\infty}^{\infty} M(\xi-y)\big(f(\xi+\alpha)-f(\xi)\big)\,d\xi\\ &= \int_{-\infty}^{\infty} M(\xi-y)\,g(\xi)\,d\xi.\end{aligned} \tag{8.03}$$

[28] Schmidt (1), (2). Throughout this section, the author has been strongly influenced by the methods of Vijayaraghavan and Szász.

Thus if $M(x)$ is subject to the conditions in the hypothesis to Theorem XII, if it is bounded above or below, and (7.07) is valid, and (7.14) bounded, whether (6.02) is true or not, it follows from Theorem XII itself on replacing $f(x)$ by $\frac{1}{\alpha}\int_{-\infty}^{x} g(\xi)\,d\xi$ that

$$
\begin{aligned}
B &= \lim_{y\to\infty} \frac{1}{\alpha}\int_{y}^{y+\alpha} \frac{dx}{\alpha}\int_{-\infty}^{x} g(\xi)\,d\xi \\
&= \lim_{y\to\infty} \frac{1}{\alpha}\int_{y}^{y+\alpha} \frac{dx}{\alpha}\int_{x}^{x+\alpha} f(\xi)\,d\xi \qquad (8.04)\\
&= \lim_{y\to\infty} \frac{1}{\alpha^2}\int_{-\alpha}^{\alpha} f(\xi+\alpha+y)\,(\alpha-|\xi|)\,d\xi.
\end{aligned}
$$

Let us now introduce Schmidt's notion of a "slowly decreasing" function. Let $f(u)$ have the property that when u and v run through a sequence of pairs of values for which

$$v \geqq u, \qquad v-u\to 0 \tag{8.05}$$

then

$$\overline{\lim_{u\to\infty}}\,(f(v)-f(u)) \geqq 0; \tag{8.06}$$

we shall call $f(u)$ *slowly decreasing*. (Schmidt treats sequences instead of functions, and on another scale, but the difference is unimportant.) Schmidt shows that we may write

$$f(u) = f_1(u)+f_2(u) \tag{8.07}$$

where $f_2(u)$ is monotone increasing, and $f_1(u)$ satisfies the condition that whenever (8.05) is fulfilled, then

$$\lim_{u\to\infty}\,(f_1(v)-f_1(u)) = 0. \tag{8.08}$$

He proves that there exists a function $T(\alpha)$ such that

$$\overline{\lim_{\substack{u\to\infty\\ u\leqq v\leqq u+\alpha}}}\,|f_1(v)-f_1(u)| = T(\alpha) \tag{8.09}$$

and

$$\lim_{\alpha\to 0} T(\alpha) = 0. \tag{8.10}$$

He proves that a number T exists, such that

$$|f_1(v)-f_1(u)| \leqq T\cdot(v-u) \qquad (v>u+\alpha). \tag{8.11}$$

From this, it readily follows that if f is bounded near $-\infty$, (8.02) is valid and $g(x)$ is bounded below. Furthermore,

$$\overline{\lim_{x\to\infty}}\left(f(x)-\frac{1}{\alpha^2}\int_{-\alpha}^{\alpha} f(\xi+\alpha+x)\,(\alpha-|\xi|)\,d\xi\right) \leqq T(2\alpha) \tag{8.12}$$

and

$$\overline{\lim_{x\to\infty}}\left(-\frac{1}{\alpha^2}\int_{-\alpha}^{\alpha} f(\xi+\alpha+x)\,(\alpha-|\,\xi\,|)\,d\xi+f(x+2\alpha)\right)\geqq -T(2\alpha); \tag{8.13}$$

going to the limit

$$\overline{\lim_{x\to\infty}} f(x)\geqq -T(2\alpha)+B \tag{8.14}$$

and

$$\overline{\lim_{x\to\infty}} f(x)\leqq B+T(2\alpha). \tag{8.15}$$

Then by (8.10), it follows that

$$\lim_{x\to\infty} f(x) = B. \tag{8.16}$$

This yields

THEOREM XIV. *In the hypothesis of Theorem* XII, *condition* (6.02) *may be replaced by the assumption that* $f(x)$ *is of limited total variation over any finite range, and is slowly decreasing, and the conclusion* (7.12) *may be replaced by* (8.16). Again, Theorem XIII becomes

THEOREM XV. *In the hypothesis to Theorem* XIII, *condition* (7.16) *may be replaced by the condition that if* u *and* v *run through a sequence for which*

$$v\geqq u,\quad \frac{v}{u}\to 1 \tag{8.17}$$

then

$$\overline{\lim_{u\to\infty}}\,(F(v)-F(u))\geqq 0. \tag{8.18}$$

In the conclusion to Theorem XIII, (7.21) *may then be replaced by*

$$\lim_{x\to\infty} F(x) = B. \tag{8.161}$$

In the corollary to Theorem XIII, (7.25) *may be replaced by the hypothesis that if*

$$s_n = a_0+a_1+\cdots+a_n \tag{8.19}$$

and $p\to\infty$, $q\to\infty$ *in such a manner that*

$$q>p,\quad \frac{q}{p}\to 1 \tag{8.20}$$

then

$$\lim_{p\to\infty}\,(s_q-s_p)\geqq 0. \tag{8.21}$$

The conclusion of this corollary remains unchanged.

The hypothesis that (8.20) shall imply (8.21) is Schmidt's hypothesis concerning slowly decreasing sequences in unaltered form.

It will be noted that the proof here given for the Schmidt theorems does not essentially differ from the proofs given by Vijayaraghavan[29] and Szász[30] for theorems of these types. In both cases the transition is made to a Tauberian theorem of more standard type in which s_n is itself averaged, instead of appearing in the conclusion as the average of a unilaterally or bilaterally bounded mass distribution. In both cases, moreover, the Schmidt condition is again used to give a bilateral or unilateral estimate of the difference between s_n and its average.

CHAPTER IV.

TAUBERIAN THEOREMS AND PRIME NUMBER THEORY.

9. **Tauberian theorems and Lambert series.** One of the most important applications of Tauberian theorems is to the proof of the prime number theorem of Hadamard[31] and de la Vallée Poussin,[32] to the effect that the number of primes less than N is asymptotically $N/\log N$. The prime numbers bear an exceedingly close relation to series of the form

$$\sum a_n \frac{x^n}{1-x^n} \tag{9.01}$$

known as Lambert series, after their eighteenth century discoverer. Until recent times, however, all attempts to employ Lambert series effectively in the study of prime numbers had proved a failure, and indeed Knopp[33] has characterized one of these directions of attack as "verführerisch". Hardy and Littlewood[34] finally showed that the prime number theorem was equivalent to a Tauberian theorem concerning Lambert series, but did not succeed in establishing an autonomous proof of this theorem.

Our general Tauberian theorems suffice to furnish this autonomous proof, and indeed, the Tauberian theorem which we shall find it easiest to establish directly leads more directly to the prime number theorem than does the theorem of Hardy and Littlewood. The latter is also directly demonstrable by our methods. In both, the cardinal point in the proof is that the Riemann zeta function, $\zeta(x+iy)$, has no zeros on the line $x = 1$. The proof of this goes back to the first proofs of the prime number theorem, and has always been recognized as that property of the Riemann zeta

[29] Vijayaraghavan (1), (2).
[30] Szász (1), (2).
[31] Hadamard (1).
[32] de la Vallée Poussin (1).
[33] Knopp (3).
[34] Hardy and Littlewood (11); Hardy (6); Ananda-Rau (2).

function which is most central in the proofs of this theorem, but all earlier proofs had made some use of the behavior of the zeta function at infinity. These further properties now appear as inessential, and the non-vanishing of the function becomes the only non-elementary feature of the zeta function in question.

For example, the usual proof of the prime number theorem employs a lemma of Landau[35] to the following effect:

LANDAU'S LEMMA. *Let*

$$F(x) = \sum_{n=1}^{\infty} a_n n^{-x} \qquad [\Re(x) > 1] \tag{9.02}$$

and let

$$a_n \geqq 0 \qquad [n = 1, 2, \cdots]. \tag{9.03}$$

Let $F(x)$ when analytically continued be without singularities on $\Re(x) = 1$, except for a pole of order one at $x = 1$, with principal part $A/(x-1)$. Let there be some α for which

$$F(x) = O(|x|^{\alpha}) \tag{9.04}$$

in the right half-plane. Then

$$A = \lim_{n \to \infty} \frac{1}{n} \sum_{1}^{n} a_k. \tag{9.05}$$

In the following section, we shall show—following Ikehara— that condition (9.04) is not needed. When the lemma is used in the proof of the prime number theorem,

$$a_n = \Lambda_n \text{ (to be defined immediately)}; \qquad F(x) = -\zeta'(x)/\zeta(x) \tag{9.06}$$

and it will be seen that here too the only non-elementary property of the zeta function which we use is that it does not vanish on the line $\Re(x) = 1$.

To return to Lambert series, let $\Lambda(n)$ be the number-theoretic function determined by:

$\Lambda(p^k) = \log p$ if p is a prime and k is a positive integer;
$\Lambda(n) = 0$ if n is not of the form p^k.

Let $0 \leqq x < 1$; then

$$\sum_{1}^{\infty} \log m\, x^m = \sum_{m=1}^{\infty} x^m \sum_{n/m} \Lambda(n) = \sum_{n=1}^{\infty} \Lambda(n) \sum_{n/m} x^m = \sum_{n=1}^{\infty} \Lambda(n) \frac{x^n}{1-x^n}. \tag{9.07}$$

[35] Landau (1).

Hence

$$\begin{aligned}\sum_1^\infty \Lambda(n)\,\frac{x^n}{1-x^n} &= \sum_1^\infty \log m\,\frac{x^m-x^{m+1}}{1-x}\\ &= \frac{x}{1-x}\sum_1^\infty \log\left(1+\frac{1}{m}\right)x^m\\ &= \frac{x}{1-x}\sum_1^\infty\left[\frac{1}{m}+O\left(\frac{1}{m^2}\right)\right]x^m\\ &= \frac{x}{1-x}\left[\log\frac{1}{1-x}+\sum_1^\infty O\left(\frac{1}{m^2}\right)x^m\right].\end{aligned} \tag{9.08}$$

If we put $x = e^{-\xi}$ and multiply by ξ we have

$$\sum_1^\infty \Lambda(n)\,\frac{\xi e^{-n\xi}}{1-e^{-n\xi}} = \frac{\xi e^{-\xi}}{1-e^{-\xi}}\left[\sum_1^\infty O\left(\frac{1}{m^2}\right)e^{-m\xi}-\log(1-e^{-\xi})\right]. \tag{9.09}$$

For $\xi>0$, the derived series on both sides converge absolutely and uniformly, and we may differentiate (9.09) term by term. On multiplication by ξ this yields us

$$\begin{aligned}&\xi\sum_1^\infty \Lambda(n)\,\frac{d}{dn\xi}\left(\frac{n\xi e^{-n\xi}}{e^{-n\xi}-1}\right)\\ &\quad= \frac{\xi e^{-\xi}(1-\xi-e^{-\xi})}{(1-e^{-\xi})^2}\left[\log(1-e^{-\xi})-\sum_1^\infty O\left(\frac{1}{m^2}\right)e^{-m\xi}\right]\\ &\quad\quad+\frac{\xi^2 e^{-\xi}}{1-e^{-\xi}}\left[\frac{e^{-\xi}}{1-e^{-\xi}}-\sum_1^\infty O\left(\frac{1}{m}\right)e^{-m\xi}\right]\\ &\quad= O(\xi)\,[O(\log\xi)-O(1)]+[\xi+O(\xi^2)]\left[\frac{1}{\xi}+O(\log\xi)\right]\\ &\quad= 1+O(\xi\log\xi)\end{aligned} \tag{9.10}$$

as $\xi\to 0$. If we put

$$N_1(u) = \frac{d}{du}\left(\frac{ue^{-u}}{e^{-u}-1}\right) \tag{9.11}$$

this leads us to

$$\lim_{\xi\to 0}\xi\sum_1^\infty \Lambda(n)\,N_1(n\xi) = 1. \tag{9.12}$$

This is a particular case of (6.14), which forms part of the hypothesis to Theorem XI‴. $\Lambda(n)$ is positive, as the hypothesis of that theorem further demands. The function $N_1(u)$ may be written

$$N_1(u) = \frac{d}{du}\left(\frac{u}{1-e^u}\right) = \frac{1+(u-1)e^u}{(1-e^u)^2} = \begin{cases} O(1) & \text{as } u\to 0;\\ O(ue^{-u}) & \text{as } u\to\infty.\end{cases} \tag{9.13}$$

Thus of that part of the hypothesis of Theorem XI‴ containing N_1, which is now both M and the whole class Σ, it only remains for us to verify the boundedness of

$$\xi \sum_1^\infty \Lambda(n)\, N_1(n\xi) \tag{9.14}$$

which follows from the fact that there is a K such that for $\xi > \xi_0 > 0$

$$\begin{aligned} \left| \xi \sum_1^\infty \Lambda(n)\, N_1(n\xi) \right| &\leqq K \sum_1^\infty n(n\xi e^{-n\xi}) \\ &= K\xi^2 \frac{d^2}{d\xi^2} \frac{e^{-\xi}}{1-e^{-\xi}} \to 0 \quad \text{as} \quad \xi \to 0 \end{aligned} \tag{9.15}$$

and from (9.12), and to verify further that

$$\int_0^\infty N_1(u)\, u^{ix}\, du \neq 0. \tag{9.16}$$

Now

$$\begin{aligned} &\int_0^\infty N_1(u)\, u^{ix}\, du \\ &= \int_0^\infty u^{ix}\, d\left(\frac{u}{1-e^u}\right) \\ &= \lim_{\lambda \to 0} \int_0^\infty u^{ix+\lambda}\, d\left(\frac{u}{1-e^u}\right) \\ &= -\lim_{\lambda \to 0} \int_0^\infty \frac{u}{1-e^u}\, d(u^{ix+\lambda}) \\ &= -\lim_{\lambda \to 0} (ix+\lambda) \int_0^\infty u^{ix+\lambda} \frac{du}{1-e^u} \qquad (9.17) \\ &= \lim_{\lambda \to 0} (ix+\lambda) \int_0^\infty u^{ix+\lambda} (e^{-u} + e^{-2u} + \cdots)\, du \\ &= \lim_{\lambda \to 0} (ix+\lambda)\, \Gamma(ix+\lambda+1)\, (1 + 2^{-(ix+\lambda+1)} + 3^{-(ix+\lambda+1)} + \cdots) \\ &= \lim_{\lambda \to 0} (ix+\lambda)\, \Gamma(ix+\lambda+1)\, \zeta(ix+\lambda+1) \\ &= ix\, \Gamma(ix+1)\, \zeta(ix+1). \end{aligned}$$

Let us take over from the theory of the Riemann zeta function the following facts:

(a) that the Riemann zeta function $\zeta(z)$ is analytic on the line with real part 1, except for a pole of the first order with principal part $1/(z-1)$;

(b) that the Riemann zeta function has no zeros with real part 1.

It then follows that

$$\int_0^\infty N_1(u)\, u^{ix}\, du \neq 0 \tag{9.18}$$

for any real x, and that

$$\int_0^\infty N_1(u)\,du = \lim_{\lambda\to 0} \lambda\,\Gamma(\lambda+1)\,\zeta(\lambda+1) = 1. \tag{9.19}$$

Thus it follows from (6.35) that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \Lambda(n) = 1. \tag{9.22}$$

Let us put $\pi(u)$ for the number of primes less than or equal to u, and let us write

$$\varpi(u) = \pi(u) + \tfrac{1}{2}\pi(u^{1/2}) + \tfrac{1}{3}\pi(u^{1/3}) + \cdots. \tag{9.23}$$

It is easy to prove by elementary means that

$$\lim_{u\to\infty} \frac{\pi(u)}{u} = 0. \tag{9.24}$$

Moreover,

$$\begin{aligned} |\varpi(u) - \pi(u)| &= \left| \tfrac{1}{2}\pi(u^{1/2}) + \cdots + \frac{\pi\left(u^{\frac{1}{[\log u/\log 2+1]}}\right)}{[\log u/\log 2]+1} \right| \\ &\leqq u^{1/2}([\log u/\log 2]+1) \qquad (9.25) \\ &= O(u^{1/2}\log u). \end{aligned}$$

Thus

$$\lim_{u\to\infty} \frac{\varpi(u)}{u} = 0. \tag{9.26}$$

We may write (9.22) in the form

$$\lim_{N\to\infty} \frac{1}{N} \int_0^N \log u\, d\varpi(u) = 1. \tag{9.27}$$

On integrating by parts, we see that

$$\lim_{N\to\infty} \left\{ \frac{\varpi(N)\log N}{N} - \frac{1}{N}\int_0^N \frac{\varpi(u)\,du}{u} \right\} = 1. \tag{9.28}$$

Now

$$\lim_{N\to\infty} \frac{1}{N} \int_0^N \frac{\varpi(u)}{u}\,du = 0 \tag{9.29}$$

because of (9.26). Thus

$$\lim_{N\to\infty} \frac{\varpi(N)\log N}{N} = 1 \tag{9.30}$$

or

$$\varpi(N) \sim \frac{N}{\log N}. \tag{9.31}$$

From (9.25) it follows that

$$\pi(N) \sim \frac{N}{\log N}. \tag{9.32}$$

This is the famous prime number theorem of de la Valleé Poussin and Hadamard.

10. **Ikehara's Theorem.** The Landau Theorem (XVI) received several successive generalizations at the hands of Landau himself, and of Hardy and Littlewood, perhaps the most general of which was indicated by Hardy and Littlewood[36] to be the following:

THEOREM XVI. *Let:*

(i) *the series* $\Sigma a_n \lambda_n^{-s}$ *be absolutely convergent for* $\Re(s) > \sigma_0 > 0$;

(ii) *the function* $F(s)$ *defined by the series be regular for* $\Re(s) > c$ *where* $0 < c \leqq \sigma_0$ *and continuous for* $\Re(s) \geqq c$, *except for a simple pole with residue* g *at* $s = c$;

(iii)

$$F(s) = O(e^{C|t|}) \tag{10.01}$$

for some finite C, *uniformly for* $\sigma \geqq c$;

(iv)

$$\frac{\lambda_n}{\lambda_{n-1}} \to 1; \tag{10.02}$$

(v) a_n *be real, and satisfy one of the inequalities*

$$a_n > -K\lambda_n^{c-1}(\lambda_n - \lambda_{n-1}), \quad a_n < K\lambda_n^{c-1}(\lambda_n - \lambda_{n-1}) \tag{10.03}$$

or complex, and of the form

$$O\{\lambda_n^{c-1}(\lambda_n - \lambda_{n-1})\}. \tag{10.04}$$

Then

$$A_n = a_1 + a_2 + \cdots + a_n \sim \frac{g\lambda_n^c}{c}. \tag{10.05}$$

The vital change between the Landau and the Hardy-Littlewood theorem is the looser form of (iii), which replaces a restriction of the form

$$F(s) = O(|t|^k). \tag{9.04}$$

Both these restrictions are inessential, and the true theorem is that of Ikehara[37], which reads as follows:

THEOREM XVII. *Let* $\alpha(x)$ *be a monotone increasing function, and let*

$$\int_{1+0}^{\infty} x^{-u}\, d\alpha(x) = f(u) \tag{10.06}$$

[36] Hardy and Littlewood (8).

[37] Ikehara (1).

converge for $\Re(u)>1$. *Let*

$$f(u)-\frac{A}{u-1}=g(u) \tag{10.07}$$

converge uniformly to a finite limit as

$$\Re(u)\to 1 \tag{10.08}$$

over any finite interval of the line $\Re(u)=1$. *Then*

$$A=\lim_{N\to\infty}\frac{\alpha(N)}{N}. \tag{10.09}$$

To prove this, let us put

$$\beta(\xi)=\alpha(e^{\xi})\,e^{-\xi}+\int_0^{\xi}e^{-\xi}\,\alpha(e^{\xi})\,d\xi-A\xi. \qquad (\xi>0). \tag{10.10}$$

Thus

$$d\beta(\xi)=e^{-\xi}\,d\alpha(e^{\xi})-A\,d\xi. \tag{10.11}$$

Let us assume—what is no essential restriction—that

$$\beta(x)=\beta(+0), \qquad (-\infty\leqq x\leqq 0). \tag{10.12}$$

Then (10.07) becomes

$$g(u)=\int_{-\infty}^{\infty}e^{(1-u)\xi}\,d\beta(\xi), \qquad [\Re(u)>1]. \tag{10.13}$$

What we wish to prove is that

$$\lim_{\eta\to\infty}\int_{-\infty}^{\eta}e^{\xi-\eta}\,d\beta(\xi)=0 \tag{10.14}$$

which is equivalent to (10.09).

If $\varepsilon>0$ and η is real, we have

$$\begin{aligned}\int_{-B}^{B}\left(1-\frac{|u|}{B}\right)g(iu+\varepsilon+1)\,e^{iu\eta}\,du\\ =-\int_{-B}^{B}\left(1-\frac{|u|}{B}\right)du\int_{-\infty}^{\infty}e^{iu(\eta-\xi)}\,e^{-\varepsilon\xi}\,d\beta(\xi).\end{aligned} \tag{10.15}$$

As this double integral is absolutely convergent, it becomes

$$\begin{aligned}\int_{-\infty}^{\infty}e^{-\varepsilon\xi}\,d\beta(\xi)\int_{-B}^{B}\left(1-\frac{|u|}{B}\right)e^{iu(\eta-\xi)}\,du\\ =-\int_{-\infty}^{\infty}\frac{2(\cos B(\eta-\xi)-1)}{B(\eta-\xi)^2}\,e^{-\varepsilon\xi}\,d\beta(\xi).\end{aligned} \tag{10.16}$$

Now,

$$\begin{aligned}\lim_{\varepsilon\to 0}\int_{-B}^{B}\left(1-\frac{|u|}{B}\right)g(iu+\varepsilon+1)\,e^{iu\eta}\,d\eta\\ =\int_{-B}^{B}\left(1-\frac{|u|}{B}\right)g(iu+1)\,e^{iu\eta}\,d\eta.\end{aligned} \tag{10.17}$$

Thus

$$\lim_{\varepsilon \to 0} \int_{-\infty}^{\infty} \frac{2(\cos B(\eta - \xi) - 1)\, e^{-\varepsilon \xi}}{B(\eta - \xi)^2}\, d\beta(\xi) \tag{10.18}$$

exists. Remembering that

$$\lim_{\varepsilon \to 0} \int_{0}^{\infty} \frac{2(\cos B(\eta - \xi) - 1)}{B(\eta - \xi)^2}\, e^{-\varepsilon \xi}\, d\xi = \int_{0}^{\infty} \frac{2(\cos B(\eta - \xi) - 1)}{B(\eta - \xi)^2}\, d\xi, \tag{10.19}$$

and that

$$\frac{2(\cos B(\eta - \xi) - 1)}{B(\eta - \xi)^2}\, e^{(-1-\varepsilon)\xi}\, d\alpha(e^{\xi}) \tag{10.20}$$

is a non-positive integrand, we have, by a theorem due to Bray[38] and fundamental in the theory of the Stieltjes integral

$$\lim_{\varepsilon \to 0} \int_{0}^{\infty} \frac{2(\cos B(\eta - \xi) - 1)}{B(\eta - \xi)^2}\, e^{(-1-\varepsilon)\xi}\, d\alpha(e^{\xi}) = \int_{0}^{\infty} \frac{2(\cos B(\eta - \xi) - 1)}{B(\eta - \xi)^2}\, e^{-\xi}\, d\alpha(e^{\xi}) \tag{10.21}$$

and hence

$$\lim_{\varepsilon \to 0} \int_{-\infty}^{\infty} \frac{2(\cos B(\eta - \xi) - 1)}{B(\eta - \xi)^2}\, e^{-\varepsilon \xi}\, d\beta(\xi) = \int_{-\infty}^{\infty} \frac{2(\cos B(\eta - \xi) - 1)}{B(\eta - \xi)^2}\, d\beta(\xi). \tag{10.22}$$

Moreover,

$$-\lim_{\eta \to \infty} \int_{-\infty}^{\infty} \frac{2(\cos B(\eta - \xi) - 1)}{B(\eta - \xi)^2}\, d\beta(\xi) = \lim_{\eta \to \infty} \int_{-B}^{B} \left(1 - \frac{|u|}{B}\right) g(iu + 1)\, e^{iu\eta}\, du = 0 \tag{10.23}$$

because $\left(1 - \frac{|u|}{B}\right) g(iu+1)$ is summable over $(-B, B)$. Similarly

$$\int_{-\infty}^{\infty} \frac{2(\cos B(\eta - \xi) - 1)}{B(\eta - \xi)^2}\, d\beta(\xi) \tag{10.24}$$

may be proved to be bounded because of (10.23) and (10.12).

We know that

$$\int_{n}^{n+1} d\beta(\xi) > -A \tag{10.25}$$

[38] Bray (1).

and that all the other conditions of Theorem XI are satisfied, with the possible exception of the non-vanishing of (6.15). To see that this is also satisfied, we need only reflect that

$$\int_{-\infty}^{\infty} \frac{2(\cos B\eta - 1)}{B\eta^2} e^{iu\eta}\, d\eta = \frac{1}{2\pi}\left(1 - \frac{|u|}{B}\right) \qquad [|u| < B]. \tag{10.26}$$

Thus there are no zeros common to all these functions for all values of B, and the non-vanishing of (6.15) follows.

It will be observed that the full force of our Tauberian method is scarcely needed for this theorem. In Theorem XI, which is the critical part of the proof, the difficulty of proof is considerably lessened if $K_1(x)$ assumes such a special form as $2(\cos Bx - 1)$.

As a corollary of Theorem XVII, we may prove:

THEOREM XVIII. *Let $\gamma(x)$ be a monotone increasing function, and let*

$$\int_{1+0}^{\infty} x^{-u}\, d\gamma(x) = \varphi(u) \tag{10.27}$$

converge for $\Re(u) > 1$. Let

$$F(u) = e^{\varphi(u)}(u-1)^A \qquad [0 < A < \tfrac{1}{3}] \tag{10.28}$$

when continued analytically, be regular for $\Re(u) = 1$, and let it not vanish for $u = 1$. Then,

$$A = \lim_{N\to\infty} \frac{1}{N} \int_{1+0}^{N} \log x\, d\gamma(x). \tag{10.29}$$

To prove this, let us put

$$\begin{cases} \int_{1+0}^{x} \log \xi\, d\gamma(\xi) = \alpha(x); \\ -\varphi'(u) = f(u). \end{cases} \tag{10.30}$$

The theorem then reduces itself to Theorem XVII, provided we can establish that (10.07) approaches a finite limit as $\Re(u) \to 1$. From the regularity of (10.28), it follows that

$$\varphi(u) + A \log(u-1) \tag{10.31}$$

is regular for $\Re(u) = 1$, except for logarithmic singularities with *negative* infinities. It is also clear that there is no singularity for $u = 1$.

Now,

$$\Re(\varphi(u)) = \int_{1+0}^{\infty} x^{-\Re(u)} \cos(\Im(u) \log x)\, d\gamma(x) \tag{10.32}$$

so that

$$\Re(\varphi(1+\varepsilon+iv)) = \int_{1+0}^{\infty} x^{-(1+\varepsilon)} \cos(v \log x)\, d\gamma(x); \tag{10.33}$$

$$\Re(\varphi(1+\varepsilon+2iv)) = \int_{1+0}^{\infty} x^{-(1+\varepsilon)} \cos(2v \log x)\, d\gamma(x); \tag{10.34}$$

and

$$\varphi(1+\varepsilon) = \int_{1+0}^{\infty} x^{-(1+\varepsilon)}\, d\gamma(x). \tag{10.35}$$

Thus

$$\begin{aligned} &3\varphi(1+\varepsilon)+4\Re(\varphi(1+\varepsilon+iv))+\Re(\varphi(1+\varepsilon+2iv)) \\ &= \int_{1+0}^{\infty} x^{-(1+\varepsilon)} (3+4\cos v \log x+\cos(2v\log x))\, d\gamma(x) \end{aligned} \tag{10.36}$$

and since

$$\begin{aligned} 3+4\cos\varphi+\cos 2\varphi &= 3+4\cos\varphi+2\cos^2\varphi-1 \\ &= 2(1+\cos\varphi)^2 \\ &\geqq 0 \end{aligned} \tag{10.37}$$

it follows that

$$3\varphi(1+\varepsilon)+4\Re(\varphi(1+\varepsilon+iv))+\Re(\varphi(1+\varepsilon+2iv)) \geqq 0, \tag{10.38}$$

or

$$\Re(\varphi(1+\varepsilon+iv)) \geqq -\tfrac{3}{4}\varphi(1+\varepsilon)-\tfrac{1}{4}\Re(\varphi(1+\varepsilon+2iv)). \tag{10.39}$$

Thus

$$\begin{aligned} &\overline{\lim_{\varepsilon\to 0}} \frac{\Re(\varphi(1+\varepsilon+iv))}{\log\varepsilon} \\ &\leqq -\tfrac{3}{4}\lim_{\varepsilon\to 0}\frac{\varphi(1+\varepsilon)}{\log\varepsilon}-\tfrac{1}{4}\overline{\lim_{\varepsilon\to 0}}\frac{\Re(\varphi(1+\varepsilon+2iv))}{\log\varepsilon} \leqq \frac{3A}{4}<1. \end{aligned} \tag{10.40}$$

On the other hand, if $1+iv$ is a logarithmic singularity of φ with a negative coefficient, it is a zero of $F(u)$ of integral order n, and

$$\overline{\lim_{\varepsilon\to 0}} \frac{\Re(\varphi(1+\varepsilon+iv))}{\log\varepsilon} = n > 1. \tag{10.41}$$

Thus $\varphi(u)+A\log(u-1)$ has no logarithmic singularities with negative coefficients, for $\Re(u)=1$, and hence is analytic throughout this whole line. Thus, by differentiation,

$$f(u)-\frac{A}{u-1} \tag{10.42}$$

is analytic on the line in question, and the finiteness of the limit of (10.07) is established.

It will be observed that our proof, which completes the demonstration of Theorem XVIII, follows closely the lines of Landau's proof[39] that the Riemann zeta function has no zeros on the line $\Re(u)=1$, and includes it

[39] Landau (2).

as a particular case. The prime number theorem itself is a particular case of Theorem XVIII. Let us put

$$\gamma(x) = \varpi(x). \tag{10.43}$$

Then

$$e^{\psi(u)} = \zeta(u) \tag{10.44}$$

and the hypothesis of Theorem XVIII is manifestly satisfied. (10.29) then becomes

$$1 = \lim_{N\to\infty} \frac{1}{N} \int_{1+0}^{N} \log x \, d\varpi(x) \tag{9.27}$$

which we have shown to be equivalent to the prime number theorem.

Thus we have repeatedly shown that the prime number theorem is basically Tauberian in nature. It might consequently be expected that the more refined theorems as to the distribution of the primes, based on Riemann's hypothesis as to the distribution of the zeros of the zeta function, and established by Hardy, Littlewood and others, might be easy to establish on a Tauberian basis. Such a formula as

$$\lim_{\xi\to 0} \xi \sum_{1}^{\infty}{}' (\Lambda(n)-1)\, n^{\alpha} \,(n^{-\alpha}\, \xi^{-\alpha}\, N_1(n\xi)) = 0 \quad \left(0<\alpha<\frac{1}{2}\right) \tag{10.45}$$

which follows from (9.10) much as does (9.12), appears to lend a certain color to this view. Here the theorem to which this seems to lead is

$$\lim_{N\to\infty} \frac{1}{N} \sum_{1}^{N} (\Lambda(n)-1)\, n^{\alpha} = 0, \tag{10.46}$$

and the condition of non-vanishing on the Fourier transform becomes the Riemann condition

$$\zeta(ix+1-\alpha) \neq 0. \tag{10.47}$$

The author considers these hopes illusory and deceptive. Let it be noted that (10.45) does not form a satisfactory hypothesis to a Tauberian theorem until we have some hold on the boundedness of the mass distribution whose integral is

$$M(u) = \sum_{1}^{[u]} (\Lambda(n)-1)\, n^{\alpha}. \tag{10.48}$$

Such information would already presuppose as much information as (10.45) can yield concerning all smaller values of α than occur in (10.46). In other words, Tauberian theorems merely transform a O into a o.

Another way of stating the same thing is to say that a Tauberian theorem always operates in the neighborhood of a single ordinate in the plane of the zeta function. This is because it depends on a division of the range of this function into near and remote parts, and because this division has validity in the theory of functions of a real variable, not in the theory of functions of a complex variable. On the other hand, the more refined properties of the distribution of primes depend on the behavior of the zeta function in the entire strip between ordinates $\frac{1}{2}$ and 1, inclusive, and can only be discussed with the aid of Cauchy's theorem.

Of course, no proof of the limitations of so vague a thing as a method has real mathematical cogency. At any time some super-Tauberian theorem may come to light and prove to be central in the utmost refinements of prime number theory. For the present, however, Tauberian theorems do not seem to lie on the main avenue of progress.

CHAPTER V.

SPECIAL APPLICATIONS OF TAUBERIAN THEOREMS.

11. **On the proof of special Tauberian theorems.** In the sequel, we shall show that the greater part of all known Tauberian theorems may be proved without great difficulty on the basis of the general theorems of the present paper. However, most of the particular theorems were proved in the first instance by entirely different methods. In individual cases, these methods are simpler and more direct than the general method here indicated. This is especially noticeable in the case of Karamata's[40] proof of the original Abel-Tauber theorem.

Being in possession of a general method, we may consider with advantage the particular methods and why they function. All Tauberian theorems of the type discussed in this paper are intimately related to the solution of an integral equation of the form

$$\int_{-\infty}^{\infty} F(y)\, G(x-y)\, dy = H(x). \tag{11.01}$$

The most direct and general method of solving such an equation is by the use of Fourier transforms. Nevertheless, there are many cases in which a repeated differentiation will reduce such an equation to a linear differential equation of finite order, and many more where the same repeated differentiation will lead to a differential equation of infinite order, but of manageable form. Thus it is appropriate in many cases to employ a technique

[40] Karamata (2), (3), (4).

of repeated differentiation, and this has been done by Hardy, Littlewood[41], and Vijayaraghavan,[42] though scarcely from an explicit consideration of the integral equations in question. So far, the successes of this method have been confined to cases where the analytic properties of the Fourier transform of F are extremely simple, and it has failed to throw any light on the Tauberian theorems of prime number theory.

The methods of Robert Schmidt[43] lie more along the lines of the present paper, in as much as he has seen the essential role played by the integral

$$\int_0^y K(x)\, x^u\, dx = R(u) \tag{11.02}$$

in the study of the kernel of a Tauberian theorem. However, he has devoted his attention to $R(u)$ for real integral arguments instead of for complex general arguments. Furthermore, Schmidt's general theorem concerns the unicity of the solution of his moment problem rather than its existence theory. As a consequence there is a wide gap between his general moment theorem and the particular Tauberian theorems which he obtains as corollaries. This gap he actually fills in in two cases, that of the Abel-Tauber theorem and that of the Borel-Tauber theorem, but he gives no general method by which it may be filled in in a new case. His actual procedure is closely allied to that of Hardy, Littlewood and Vijayaraghavan. Schmidt's chief service to the subject is in his great improvement of what may be called the auxiliary apparatus of the theory of Tauberian theorems, through his invention of the notions of "langsam abfallende Funktionen" and "gestrahlte Matrizen".

Karamata's elegant method leads to the study of the closure of a set of translations of a given function, and thus most closely approximates to that developed here. His function is

$$f(x) = e^x e^{-e^x} \tag{11.03}$$

and the problem is solved through Weierstrass' theory of polynomial approximation. The translations considered are accordingly those of the form

$$f(x + \log \mu). \tag{11.04}$$

Szász has carried further the study of this particular set of translations of a given function. This is a far more difficult study than that of the

[41] Cf. Littlewood (1).

[42] Vijayaraghavan (1), (2).

[43] Schmidt (1), (2).

closure of the complete set of translations, and here the general solution of the closure problem is not yet known to me.

12. **Examples of kernels for which Tauberian theorems hold.** Among the kernels admissible in the role of the $N(x)$ of Theorem XIII are:

(1) The Riesz kernels[44]

$$N(x) = \begin{cases} (1-x)^\lambda; & [0<x<1] \\ 0 \quad ; & [1<x] \end{cases} \qquad [\lambda > 0]. \tag{12.01}$$

(2) The Abel kernel

$$N(x) = e^{-x}; \tag{12.02}$$

(3) The kernel

$$N(x) = \frac{1}{1+x^2}, \tag{12.03}$$

corresponding to the method of summation of the series Σa_n, which gives as its partial sum the Abel average

$$\lambda \int_0^\infty f(x)\, e^{-\lambda x}\, dx \tag{12.04}$$

of the cosine series

$$f(x) = \sum a_n \cos nx. \tag{12.05}$$

In the respective cases we have

(1)

$$\int_0^\infty N(x)\, x^{z-1}\, dx = \frac{\Gamma(\lambda+1)\,\Gamma(z)}{\Gamma(\lambda+z+1)}; \tag{12.06}$$

(2)

$$\int_0^\infty N(x)\, x^{z-1}\, dx = \Gamma(z); \tag{12.07}$$

(3)

$$\int_0^\infty N(x)\, x^{z-1}\, dx = \frac{\pi}{2 \sin \frac{\pi z}{2}}. \tag{12.08}$$

None of these functions vanishes for purely imaginary values of z.

A possible $N_1(\mu)$ of Theorem XI''' is

$$N_1(\mu) = \begin{cases} (\lambda+1)\,(1-\mu)^\lambda; & [0<\mu<1] \\ 0 & [1<\mu] \end{cases} \qquad [\lambda \geqq 0]. \tag{12.09}$$

Here

$$\int_0^\infty N_1(\mu)\, \mu^{iu}\, d\mu = \frac{\Gamma(\lambda+2)\,\Gamma(iu+1)}{\Gamma(iu+\lambda+2)} \neq 0 \qquad \text{if } u \text{ is real.} \tag{12.10}$$

[44] Riesz (1).

13. **A theorem of Ramanujan.** As a lemma in our further work, we shall find it convenient to introduce a theorem formulated by Ramanujan[45] and first proved by Hardy and Titchmarsh,[46] although with a formulation somewhat different from that here given. The theorem is intrinsically interesting, and is perhaps worth presenting in some detail.

Let $f(x)$ be a function of L_2, defined over $(0, \infty)$. Then $e^{x/2} f(e^x)$ will also belong to L_2 over $(-\infty, \infty)$; for

$$\int_{-\infty}^{\infty} |e^{x/2} f(e^x)|^2 \, dx = \int_{-0}^{\infty} |f(x)|^2 \, dx. \tag{13.01}$$

The Fourier transform of $e^{x/2} f(e^x)$ will be

$$\begin{aligned} h(u) &= \sqrt{\frac{1}{2\pi}} \operatorname*{l.i.m.}_{A\to\infty} \int_{-A}^{A} e^{iux} e^{x/2} f(e^x) \, dx \\ &= \frac{1}{\sqrt{2\pi}} \operatorname*{l.i.m.}_{\varepsilon\to 0} \int_{\varepsilon}^{1/\varepsilon} f(x) \, x^{iu-1/2} \, dx. \end{aligned} \tag{13.02}$$

The sine transform of $f(x)$ will be

$$g_1(y) = \sqrt{\frac{2}{\pi}} \operatorname*{l.i.m.}_{B\to\infty} \int_0^B f(x) \sin xy \, dx, \tag{13.03}$$

and its cosine transform,

$$g_2(y) = \sqrt{\frac{2}{\pi}} \operatorname*{l.i.m.}_{B\to\infty} \int_0^B f(x) \cos xy \, dx. \tag{13.04}$$

Let us put

$$k_1(u) = \frac{1}{\sqrt{2\pi}} \operatorname*{l.i.m.}_{\varepsilon\to 0} \int_{\varepsilon}^{1/\varepsilon} g_1(y) \, y^{iu-1/2} \, dy, \tag{13.05}$$

and

$$k_2(u) = \frac{1}{\sqrt{2\pi}} \operatorname*{l.i.m.}_{\varepsilon\to 0} \int_{\varepsilon}^{1/\varepsilon} g_2(y) \, y^{iu-1/2} \, dy. \tag{13.06}$$

We have

$$\begin{aligned} \frac{1}{z} \int_0^z g_1(y) \, dy &= \sqrt{\frac{2}{\pi}} \int_0^{\infty} f(x) \, dx \frac{1}{z} \int_0^z \sin xy \, dy \\ &= \sqrt{\frac{2}{\pi}} \int_0^{\infty} f(x) \frac{1-\cos xz}{xz} \, dx. \end{aligned} \tag{13.07}$$

Similarly

$$\frac{1}{z} \int_0^z g_2(y) \, dy = \sqrt{\frac{2}{\pi}} \int_0^{\infty} f(x) \frac{\sin xz}{xz} \, dx. \tag{13.08}$$

[45] Hardy and Littlewood (8).

[46] Hardy and Titchmarsh (1).

However, on an exponential transformation, these become

$$(13.09)\quad \int_{-\infty}^{\zeta} e^{\frac{\eta-\xi}{2}} (e^{\eta/2} g_2(e^{\eta}))\, d\eta = \sqrt{\frac{2}{\pi}} \int_{-\infty}^{\infty} \sin(e^{\xi+\zeta})\, e^{-\frac{\xi+\zeta}{2}} (e^{\xi/2} f(e^{\xi}))\, d\xi,$$

and

$$(13.10)\quad \int_{-\infty}^{\zeta} e^{\frac{\eta-\zeta}{2}} (e^{\eta/2} g_1(e^{\eta}))\, d\eta = \sqrt{\frac{2}{\pi}} \int_{-\infty}^{\infty} (1-\cos(e^{\xi+\zeta}))\, e^{-\frac{\xi+\zeta}{2}} (e^{\xi/2} f(e^{\xi}))\, d\xi.$$

If we now make a Fourier transformation and make use of the Parseval theorem, it appears that

$$(13.11)\quad k_1(u) \int_0^{\infty} e^{-\eta/2}\, e^{iu\eta}\, d\eta = \sqrt{\frac{2}{\pi}}\, h(-u) \int_{-\infty}^{\infty} (1-\cos e^{\xi})\, e^{-\xi/2}\, e^{iu\xi}\, d\xi,$$

and

$$(13.12)\quad k_2(u) \int_0^{\infty} e^{-\eta/2}\, e^{iu\eta}\, d\eta = \sqrt{\frac{2}{\pi}}\, h(-u) \int_{-\infty}^{\infty} \sin e^{\xi}\, e^{-\xi/2}\, e^{iu\xi}\, d\xi.$$

Thus

$$(13.13)\quad \begin{aligned} \frac{k_1(u)}{h(-u)} &= \frac{\sqrt{\frac{2}{\pi}} \int_0^{\infty} (1-\cos x)\, x^{iu-3/2}\, dx}{\int_0^{\infty} e^{-\eta(1/2-iu)}\, d\eta} \\ &= \sqrt{\frac{2}{\pi}} \sin\left(iu+\frac{1}{2}\right) \frac{\pi}{2}\, \Gamma\left(iu+\frac{1}{2}\right). \end{aligned}$$

Here we make use of the formulae:

$$(13.131)\quad \int_0^{\infty} x^{\nu} \sin x\, dx = \Gamma(\nu+1) \cos\frac{\nu\pi}{2} \qquad [-1 \geqq \Re(\nu) > -2]$$

and

$$(13.132)\quad \int_0^{\infty} x^{\nu} (1-\cos x)\, dx = \Gamma(\nu+1) \sin\frac{\nu\pi}{2} \qquad [-1 > \Re(\nu) > -3].$$

Similarly

$$(13.14)\quad \frac{k_2(u)}{h(-u)} = \sqrt{\frac{2}{\pi}} \cos\left(iu+\frac{1}{2}\right) \frac{\pi}{2}\, \Gamma\left(iu+\frac{1}{2}\right).$$

The duality of the relation between $k_1(u)$ or $k_2(u)$ and $h(u)$ is shown by the familiar formulae of the gamma function,

$$(13.15)\quad \begin{aligned} &\frac{2}{\pi} \sin\left(iu+\frac{1}{2}\right)\frac{\pi}{2} \sin\left(-iu+\frac{1}{2}\right)\frac{\pi}{2}\, \Gamma\left(iu+\frac{1}{2}\right) \Gamma\left(-iu+\frac{1}{2}\right) \\ &= \frac{2}{\pi} \cos\left(iu+\frac{1}{2}\right)\frac{\pi}{2} \cos\left(-iu+\frac{1}{2}\right)\frac{\pi}{2}\, \Gamma\left(iu+\frac{1}{2}\right) \Gamma\left(-iu+\frac{1}{2}\right) \\ &= \frac{1}{\pi} \cos iu\pi\, \Gamma\left(iu+\frac{1}{2}\right) \Gamma\left(-iu+\frac{1}{2}\right) = 1. \end{aligned}$$

Thus

$$\left|\frac{k_2(u)}{h(-u)}\right| = 1, \quad \left|\frac{k_1(u)}{h(-u)}\right| = 1, \tag{13.16}$$

and the real zeros of $k_1(u)$, $k_2(u)$, and $h(-u)$ are the same. In other words, if the integrals in question exist, the zeros of

$$\left\{\begin{array}{l} \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_0^\infty f(x)\,x^{-iu-1/2}\,dx, \\ \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_0^\infty g_1(y)\,y^{iu-1/2}\,dy, \\ \text{and } \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_0^\infty g_2(y)\,y^{iu-1/2}\,dy, \end{array}\right. \tag{13.17}$$

are the same. This is a very valuable way of determining new kernels whose Fourier transforms (on a logarithmic scale) have no zeros.

14. **The summation of trigonometrical developments.** Let $f(x)$ be an even function of class L_2, and, as above, let

$$g_2(y) = \sqrt{\frac{2}{\pi}}\,\underset{B\to\infty}{\text{l. i. m.}}\int_0^B f(x)\cos xy\,dx. \tag{13.04}$$

Let $K(x)$ belong to L_2 over $(0, \infty)$, and let us form

$$x\int_0^\infty g_2(w)\,K(wx)\,dw. \tag{14.01}$$

This will equal

$$\frac{1}{x}\int_0^\infty z f(z)\frac{x}{z}\,k\left(\frac{z}{x}\right)dz, \tag{14.02}$$

where

$$k(z) = \sqrt{\frac{2}{\pi}}\,\underset{B\to\infty}{\text{l. i. m.}}\int_0^B K(w)\cos wz\,dw; \tag{14.03}$$

for the cosine transform of $xK(wx)$ is $k\left(\frac{z}{x}\right)$. Thus an average of $g_2(w)$ at infinity will appear formally as an average of $zf(z)$ about the origin, and an average of $g_2(w)$ about the origin will appear as an average of $zf(z)$ at infinity. The kernels in the two cases will be $K(x)$ and $\frac{1}{x}k(x)$. Now by § 13, we have formally

$$\left|\sqrt{\frac{1}{2\pi}}\int_0^\infty \frac{k(x)}{x}x^\lambda\,dx\right| = \left|\sqrt{\frac{1}{2\pi}}\int_0^\infty K(x)\,x^{-\lambda}\,dx\right|, \tag{14.04}$$

when $\Re(\lambda) = \frac{1}{2}$. If

$$(14.05) \qquad \sqrt{\frac{1}{2\pi}} \int_0^\infty \frac{k(x)}{x} x^\lambda \, dx \quad \text{and} \quad \sqrt{\frac{1}{2\pi}} \int_0^\infty K(x) x^{-\lambda} \, dx$$

converge for $0 \leqq \Re(\lambda) \leqq \frac{1}{2}$, it then follows that

$$(14.06) \qquad \sqrt{\frac{1}{2\pi}} \int_0^\infty \frac{k(x)}{x} x^{iu} \, dx \neq 0 \qquad [-\infty < u < \infty]$$

when, and only when,

$$(14.07) \qquad \frac{1}{\sqrt{2\pi}} \int_0^\infty K(x) x^{iu} \, dx \neq 0 \qquad [-\infty < u < \infty].$$

Thus if $k(x)/x$ in the role of $N_1(x)$, and $N_2(x)$ satisfy the other conditions specified in the hypothesis of Theorem XI, and (14.07) holds, it is possible to infer from

$$(14.08) \qquad \lim_{\substack{x \to 0 \\ \infty}} x \int_0^\infty g_2(w) K(wx) \, dw = A \int_0^\infty K(w) \, dw$$

to the conclusion that

$$(14.09) \qquad \lim_{\substack{x \to 0 \\ \infty}} x \int_0^\infty z f(z) N_2(zx) \, dz = A \int_0^\infty N_2(z) \, dz.$$

The condition that f belongs to L_2 may in many cases be very considerably altered and relaxed.

It is possible to read the relation between (14.01) and (14.02) in the reverse direction, and to infer from

$$(14.10) \qquad \lim_{x \to \infty} \int_0^\infty f(z) k\left(\frac{z}{x}\right) dz = Ak(0)$$

and other conditions completing the hypothesis of Theorem XI, to

$$(14.11) \qquad \lim_{x \to 0} x \int_0^\infty g_2(w) K_2(wx) \, dw = A \int_0^\infty K_2(w) \, dw.$$

The condition that

$$(14.12) \qquad \frac{1}{\sqrt{2\pi}} \int_0^\infty k(x) x^{s-1} \, dx$$

should determine a function over a strip to the right of the origin, which, when continued has no zeros over the imaginary axis, becomes the same condition for

$$(14.13) \qquad \frac{1}{\sqrt{2\pi}} \int_0^\infty K(x) x^{-s} \, dx.$$

We thus arrive at a whole class of theorems relating the partial sum of a Fourier development with the average of the function represented about some point. For example, it is easy to prove:

THEOREM XVIII. *If $f(x)$ is a non-negative function of class L_2 over $(-\pi, \pi)$, if it is summable at any point by a Riesz mean of any positive order, or by an Abel mean, to any value A, it is so summable to A by Riesz means of all positive orders, and by an Abel mean. A necessary and sufficient condition that this should take place at a given point x is that*

$$\lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_{x-\varepsilon}^{x+\varepsilon} f(x)\, dx = A. \tag{14.14}$$

This theorem is due to Hardy and Littlewood.

15. **Young's criterion for the convergence of a Fourier series.** We now come to a region in which S. B. Littauer has done work, which constitutes the theorems proved in the present section. Young has proved the following theorem: *for the Fourier series of the integrable function $f(u)$ to converge to s for $u = x$, it is sufficient that*

$$\varphi(t) \to 0 \tag{15.01}$$

and that

$$\int_0^t |\, d(u\,\varphi(u))\,| = O(t) \tag{15.02}$$

for small t, where

$$\varphi(t) = \frac{1}{2}\,\{f(x+t)+f(x-t)-2s\}. \tag{15.03}$$

Further work on this theorem has been done by Young, Pollard, Hardy and Littlewood. The chief theorem to which Hardy and Littlewood came was the following:

THEOREM XIX. *A necessary and sufficient condition that the Fourier series of the integrable function $f(u)$ be summable $(C, -1+\delta)$ for any positive δ is that*

$$\varphi_1(t) = o(t) \tag{15.04}$$

provided that

$$\int_0^t |\, d(u\,\varphi(u))\,| = O(t), \qquad [0 \leqq t < \pi] \tag{15.05}$$

where

$$\varphi_1(t) = \int_0^t \varphi(u)\, du. \tag{15.06}$$

This theorem in its original form is not directly adaptable to proof by the methods of this paper, inasmuch as Cesàro summation of fractional

order does not depend on a kernel of the form $N(nx)$. On the other hand, if the Cesàro summation is replaced by Riesz summation, the theorem is reduced to a particular case of theorems already proved. We shall here content ourselves with proving this related theorem, as applied to Fourier integrals rather than series, in the case where $f(x)$ differs from zero only over a finite range.[47] Let us write

$$t\,\varphi(t) = \psi(t). \tag{15.07}$$

Then (15.05) becomes

$$\int_0^t |\,d\,\psi(t)\,| = O(t), \qquad [0 \leqq t < \pi] \tag{15.08}$$

which yields the boundedness of

$$\int_\mu^{2\mu} \lambda^{-1} |\,d\,\psi(\lambda)\,| - \int_\mu^{2\mu} \lambda^{-1}\,d\,\psi(\lambda). \tag{15.09}$$

Furthermore

$$\begin{aligned} \varphi_1(t) &= \int_0^t \varphi(u)\,du = \int_0^t \psi(u)\,\frac{du}{u} \\ &= \int_0^t \frac{du}{u} \int_0^u d\,\psi(v) = \int_0^t d\,\psi(v) \int_v^t \frac{du}{u} \\ &= \int_0^t \log\frac{t}{v}\,d\,\psi(v). \end{aligned} \tag{15.10}$$

Thus (15.04) becomes

$$\lim_{t\to 0} \frac{1}{t} \int_0^t \log\frac{t}{v}\,d\,\psi(v) = 0. \tag{15.11}$$

We have

$$\int_0^1 \log\lambda\ \lambda^{iu}\,d\lambda = -(iu+1)^{-2} \neq 0. \tag{15.12}$$

This is similar in form to (6.15).

The Riesz sum of kth order of the Fourier series of $f(x) - s$ for $x = 0$ is

$$\begin{aligned} &\lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_0^\infty \frac{\varphi(t)\,dt}{\Gamma(1+k)} \int_0^1 (1-\lambda)^k \cos\frac{\lambda t}{\varepsilon}\,d\lambda \\ &= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_0^\infty \frac{dt}{t\Gamma(1+k)} \int_0^t d\,\psi(u) \int_0^1 (1-\lambda)^k \cos\frac{\lambda t}{\varepsilon}\,d\lambda \\ &= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_0^\infty d\,\psi(u) \int_u^\infty \frac{dt}{t\Gamma(1+k)} \int_0^1 (1-\lambda)^k \cos\frac{\lambda t}{\varepsilon}\,d\lambda \\ &= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_0^\infty d\,\psi(u) \int_{u/\varepsilon}^\infty \frac{dt}{t\Gamma(k+1)} \int_0^1 (1-\lambda)^k \cos\lambda t\,d\lambda. \end{aligned} \tag{15.13}$$

[47] It is possible to make the transition to Theorem XIX, but several of the steps require some consideration. In particular, the equivalence of Riesz and Cesàro summation for orders > -1 was proved by Riesz (Proc. Lond. Math. Soc., 2 (22), 412-419 (1924).

Here if $k > -1$,

$$\int_0^\infty \left| \int_0^1 (1-\lambda)^k \cos \lambda\tau \, d\lambda \right| d\tau \tag{15.14}$$

is finite, and the inversion of integration in (15.13) is hence applied to an absolutely convergent integral, in view of (15.08).

We have

$$\begin{aligned}
&\int_0^\infty \omega^{iu} \, d\omega \int_\omega^\infty \frac{d\tau}{\tau \Gamma(1+k)} \int_0^1 (1-\lambda)^k \cos \lambda\tau \, d\lambda \\
&= \frac{1}{iu+1} \int_0^\infty \frac{\omega^{iu}}{\Gamma(1+k)} \, d\omega \int_0^1 (1-\lambda)^k \cos \lambda\omega \, d\lambda \\
&= \frac{1}{iu+1} \int_0^\infty \frac{\omega^{iu}}{\Gamma(1+k)} \, d\omega \int_{\lambda=0}^{\lambda=1} (1-\lambda)^k \frac{d \sin \lambda\omega}{\omega} \\
&= \frac{1}{(iu+1)\Gamma(k)} \int_0^\infty \omega^{iu-1} \, d\omega \int_0^1 (1-\lambda)^{k-1} \sin \lambda\omega \, d\lambda \\
&= \frac{1}{(iu+1)\Gamma(k)} \int_0^1 (1-\lambda)^{k-1} \, d\lambda \int_0^\infty \omega^{iu-1} \sin \lambda\omega \, d\omega \\
&= \frac{1}{(iu+1)\Gamma(k)} \int_0^1 (1-\lambda)^{k-1} \lambda^{-iu} \, d\lambda \int_0^\infty z^{iu-1} \sin z \, dz \qquad (15.15) \\
&= \frac{1}{(iu+1)\Gamma(k)} \frac{\Gamma(k)\Gamma(1-iu)}{\Gamma(k+1-iu)} \Gamma(iu) \cos (iu-1)\frac{\pi}{2} \\
&= \frac{\Gamma(1-iu)\Gamma(iu)}{(iu+1)\Gamma(k+1-iu)} \sin \frac{\pi iu}{2} \\
&= \frac{\pi \sin \dfrac{\pi iu}{2}}{(iu+1)\Gamma(k+1-iu) \sin \pi iu} \\
&= \frac{\pi}{2(iu+1)\Gamma(k+1-iu) \cos \dfrac{\pi iu}{2}} \\
&\neq 0.
\end{aligned}$$

In proving this, we have made use of (13.131) and (13.132).

The other conditions of the hypothesis of Theorem XI′ are readily proved to be satisfied, and we have already shown that if (15.05) is satisfied, (15.11) *and*

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^\infty \frac{\varphi(t)\, dt}{\Gamma(1+k)} \int_0^1 (1-\lambda)^k \cos \frac{\lambda t}{\varepsilon} \, d\lambda = 0, \tag{15.16}$$

are equivalent. Thus if (15.05) *is satisfied,* (15.04) *is completely equivalent to the statement that the Fourier series of* $f(x)$ *is summable to* s *by Riesz (or Cesàro) sums of any given order exceeding* -1.

16. **Tauberian theorems and asymptotic series.** Certain asymptotic problems arise in the discussion of the behavior of an integral

$$\int_a^b (g(t))^x \Phi(t)\, dt \tag{16.01}$$

for large values of x. By a change of variable, this problem may be reduced to the consideration of

$$\int_0^\infty e^{-xt} \Phi(t)\, dt. \tag{16.02}$$

In particular, let us discuss the situation which arises when $\nu > 0$ and

$$\lim_{x\to\infty} x^\nu \int_0^\infty e^{-xt} \Phi(t)\, dt = A, \tag{16.03}$$

which we may write

$$\lim_{x\to\infty} x \int_0^\infty (xt)^{\nu-1} e^{-xt} \frac{\Phi(t)}{t^{\nu-1}}\, dt = A. \tag{16.04}$$

If

$$x \int_0^\infty (xt)^{\nu-1} e^{-xt} \frac{\Phi(t)}{t^{\nu-1}}\, dt \tag{16.05}$$

is bounded and $\Phi(t)$ is positive, it will follow by Theorem XI″ that

$$\lim_{x\to\infty} x \int_0^{1/x} \frac{\Phi(t)}{t^{\nu-1}}\, dt = \frac{A}{\int_0^\infty t^{\nu-1} e^{-t}\, dt} = \frac{A}{\Gamma(\nu)}, \tag{16.06}$$

or that

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_0^\varepsilon \frac{\Phi(t)}{t^{\nu-1}}\, dt = \frac{A}{\Gamma(\nu)}. \tag{16.07}$$

Again, it will follow that

$$\lim_{x\to\infty} x \int_{1/x}^\infty \frac{\Phi(t)}{t^{\nu-1}} \frac{dt}{x^2 t^2} = \frac{A}{\Gamma(\nu)}, \tag{16.08}$$

which yields

$$\lim_{\varepsilon\to 0} \varepsilon \int_0^\varepsilon t^{\nu-1} \Phi\left(\frac{1}{t}\right) dt = \frac{A}{\Gamma(\nu)}. \tag{16.09}$$

The condition that

$$\Gamma(iu+\nu) = \int_0^\infty e^{-t} t^{iu+\nu-1}\, dt \neq 0 \tag{16.10}$$

is obviously satisfied. Indeed the Tauberian theorems of this section (which are due to Ikehara[48]) differ only from those of the last section in that x tends to infinity instead of to 0.

If $\frac{\Phi(t)}{t^{\nu-1}}$ is bounded and tends to B at the origin and $\nu>0$,

$$\lim_{x\to\infty} x\int_0^\infty (xt)^{\nu-1}e^{-xt}\left[\frac{\Phi(t)}{t^{\nu-1}}-B\right]dt = 0. \tag{16.11}$$

Then

$$\lim_{x\to\infty} x\int_0^\infty (xt)^{\nu-1}e^{-xt}\frac{\Phi(t)}{t^{\nu-1}}\,dt = Bx^\nu\int_0^\infty t^{\nu-1}e^{-xt}\,dt = B\Gamma(\nu). \tag{16.12}$$

Thus

$$\lim_{x\to\infty} x^\nu\int_0^\infty e^{-xt}\Phi(t)\,dt = B\Gamma(\nu). \tag{16.13}$$

In particular, let there be a neighborhood of the origin in which $\Phi(t)$ is $\nu-1$ times differentiable, with a bounded derivative of the $(\nu-1)$st order. Let

$$\Phi(0) = 0;\quad \Phi'(0) = 0;\quad \cdots\quad \Phi^{(\nu-2)}(0) = 0;\quad \Phi^{(\nu-1)}(0) = A \tag{16.14}$$

and let $\frac{\Phi(t)}{t^{\nu-1}}$ be bounded outside the neighborhood in question. Then

$$\lim_{x\to\infty} x^\nu\int_0^\infty e^{-xt}\Phi(t)\,dt = A. \tag{16.03}$$

If now $\Phi(t)-At^{\nu-1}$ has a bounded derivative of order $\mu-1>\nu-1$ at the origin, and vanishes there with all its derivatives of order less than $\mu-1$, and $\Psi(t)$ is a function which equals $\Phi(t)-\frac{At^{\nu-1}}{\Gamma(\nu)}$ in some neighborhood of the origin, and for which $\frac{\Psi(t)}{t^{\mu-1}}$ is bounded and tends to A at the origin, we have

$$\lim_{x\to\infty} x^\mu\int_0^\infty e^{-xt}\Psi(t)\,dt = A. \tag{16.15}$$

By a repetition of this process, it is easy to show that if

$$\Phi(t) = \frac{A_1t^{\nu_1-1}}{\Gamma(\nu_1)}+\frac{A_2t^{\nu_2-1}}{\Gamma(\nu_2)}+\cdots+\frac{A_kt^{\nu_k-1}}{\Gamma(\nu_k)}+\Psi(t), \tag{16.16}$$

if $\Psi(t)$ is bounded and $(\mu-1)$ times differentiable with bounded $(\mu-1)$st derivative in some neighborhood of the origin, if

[48] In a paper not yet published.

$$(16.17)\qquad \Psi(0) = 0; \quad \Psi'(0) = 0; \quad \cdots \quad \Psi^{(\mu-1)}(0) = 0,$$

and if $\mu > \nu_1$, $\mu > \nu_2$, $\cdots$, $\mu > \nu_k$, then

$$(16.18)\qquad \int_0^\infty e^{-xt}\, \Phi(t)\, dt = A_1 x^{-\nu_1} + A_2 x^{-\nu_2} + \cdots + A_k x^{-\nu_k} + O(x^{-\mu+\varepsilon})$$

as $x \to \infty$. This is an adequate basis for the theory of asymptotic expansions of Laplace integrals, and enables Wintner's work[49] on the subject to be greatly simplified.

CHAPTER VI.

KERNELS ALMOST OF THE CLOSED CYCLE.

17. The reduction of kernels almost of the closed cycle to kernels of the closed cycle. In the present section, we shall approach very close to the work of Robert Schmidt[50] in his discussion of "gestrahlte Mittelbildungen", although our terminology will be somewhat different. Up to this point, we have been discussing means of the form

$$(17.01)\qquad \int_{-\infty}^{\infty} K_1(\xi - x)\, f(\xi)\, d\xi$$

or

$$(17.02)\qquad \int_{-\infty}^{\infty} K_1(\xi - x)\, d\varphi(\xi),$$

or means only differing from these by a change of variable from ξ to a function of ξ. Let us now turn our attention to means of the form

$$(17.03)\qquad \int_{-\infty}^{\infty} K_1(\xi, x)\, d\varphi(\xi),$$

where

$$(17.04)\qquad K_1(\xi, x) = K_1(\xi - x) + K_1^*(\xi, x)$$

where $K_1(x)$ satisfies the condition that (5.02) converges and that (5.07) does not vanish, and where

$$(17.05)\qquad \lim_{x\to\infty} \sum_{-\infty}^{\infty} \max_{n \le y \le n+1} |K_1^*(y, x)| = 0.$$

Let $K_1(x, y)$ and $K_1(x)$ be continuous, and let

$$(17.06)\qquad \int_{-\infty}^{\infty} K_1(x, y)\, d\varphi(x) \quad \text{and} \quad \int_y^{y+1} |d\varphi(x)| - \varphi(y+1) + \varphi(y)$$

[49] Wintner (1).
[50] Schmidt (2).

be bounded. Let

$$\lim_{x\to\infty}\int_{-\infty}^{\infty} K_1(\xi, x)\, d\varphi(\xi) = A\int_{-\infty}^{\infty} K_1(\xi)\, d\xi. \tag{17.061}$$

The argument of Lemma XIa will need no substantial alteration to show that there is a Q such that

$$\int_y^{y+1} |d\varphi(x)| \leqq Q. \tag{6.04}$$

It will then follow from (17.05) that

$$\lim_{x\to\infty}\int_{-\infty}^{\infty} K_1(\xi - x)\, d\varphi(\xi) = A\int_{-\infty}^{\infty} K_1(\xi)\, d\xi \tag{5.03}$$

and that

$$\int_{-\infty}^{\infty} K_1(\xi - x)\, d\varphi(\xi)$$

is bounded. Hence, by Theorem X, if K_2 is any function belonging to N_1,

$$\lim_{x\to\infty}\int_{-\infty}^{\infty} K_2(\xi - x)\, d\varphi(\xi) = A\int_{-\infty}^{\infty} K_2(\xi)\, d\xi. \tag{5.04}$$

In Chapter III, the kernel $K(\xi - x)$ is replaced by a kernel of the form

$$M(\xi + \alpha - x) - M(\xi - x) \tag{17.07}$$

where $M(x)$ is a monotone function for which

$$M(-\infty) = 1, \qquad M(x) = O(e^{-ux}) \text{ at } +\infty; \qquad [\mu > 0] \tag{7.13}$$

and for which the function

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} M(x)\, e^{xz}\, dx \tag{7.05}$$

which is defined over a strip to the right of the origin, when continued analytically on to the imaginary axis, has no zeros there. We can replace the kernel $M(\xi - x)$ by a kernel of the form

$$M(\xi, x) = M_1(\xi - x) + M_2(\xi, x) \tag{17.08}$$

where $M_1(x)$ satisfies the conditions we have already laid down for $M(x)$, and where

$$\lim_{x\to\infty}\int_0^{\infty} |dM_2(\xi, x)| = 0 \tag{17.09}$$

and

$$M_2(\xi, x) = O(e^{-u\xi}) \text{ at } \infty \text{ uniformly.} \tag{17.10}$$

Under this change, Theorem XIV still remains valid.

18. **A Tauberian theorem of Hardy and Littlewood.**[51] Hardy and Littlewood have proved the following theorem:

THEOREM XX. *Let* $f(x) = \sum a_n x^n$ *be a power series with positive coefficients, and let*

$$f(x) \sim \frac{A}{(1-x)^\alpha} \quad (\alpha > 0,\ x \to 1). \tag{18.01}$$

Then

$$\sum_1^n a_k \sim \frac{n^\alpha A}{\Gamma(\alpha+1)} \quad (n \sim \infty). \tag{18.02}$$

We may write (18.01)

$$\sum a_n e^{-n\xi} \sim A\,\xi^{-\alpha} \tag{18.03}$$

or

$$\lim_{\xi \to 0} \xi \sum \frac{a_n}{n^{\alpha-1}} (n\xi)^{\alpha-1} e^{-n\xi} = A. \tag{18.04}$$

This falls under (6.30), and

$$\int_0^\infty \xi^{\alpha-1} e^{-\xi}\, \xi^{iu}\, d\xi = \Gamma(\alpha+iu) \neq 0. \tag{18.05}$$

Since $a_n/n^{\alpha-1} > 0$,

$$\lim_{\xi \to 0} \xi \sum_1^{[1/\xi]} \frac{a_n}{n^{\alpha-1}} (n\xi)^{\alpha-1} = A \frac{\int_0^1 \xi^{\alpha-1}\, d\xi}{\int_0^\infty \xi^{\alpha-1} e^{-\xi}\, d\xi} = \frac{A}{\Gamma(\alpha+1)} \tag{18.06}$$

because of Theorem XI'''. Writing m for $1/\xi$, and letting m become infinite through integral values, we get

$$\lim_{m \to \infty} \frac{1}{m^\alpha} \sum_1^m a_n = \frac{A}{\Gamma(\alpha+1)}, \tag{18.07}$$

which is only another way of writing (18.02).

This however is not the most general theorem proved by Hardy and Littlewood in this connection. They show that if

$$f(x) \sim \frac{A}{(1-x)^\alpha}\, L\left(\frac{1}{1-x}\right), \tag{18.08}$$

where

$$L(u) = (\log u)^{\alpha_1} (\log\log u)^{\alpha_2} \cdots (\log^{(n)} u)^{\alpha_n} \tag{18.09}$$

and

$$u^\alpha L(u) \neq O(1) \text{ at } \infty, \tag{18.10}$$

then

$$\sum_1^n a_k \sim \frac{n^\alpha A\, L(n)}{\Gamma(\alpha+1)}. \tag{18.11}$$

[51] Hardy and Littlewood (4).

We may write our theorem to be proved in the form that if

$$\xi \sum_1^\infty \frac{a_n}{n^{\alpha-1}} (n\xi)^{\alpha-1} e^{-n\xi} \sim A L(\xi^{-1}), \tag{18.12}$$

then (18.11) follows. This we may again write

$$e^{-\mu} \sum_1^\infty \frac{a_n}{n^{\alpha-1}} e^{(\log n-\mu)(\alpha-1)} e^{-e^{\log n-\mu}} \sim A L(e^\mu) \tag{18.13}$$

as $\mu \to \infty$. Now, if $M(\xi)$ is a continuous function, defined over $(0, \infty)$, and asymptotic to $L(\xi)$ at ∞, we may show that:

$$\int_0^\infty e^{(u-\mu)\alpha} e^{-e^{(u-\mu)}} M(e^u)\, du \sim M(e^\mu) \int_{-\infty}^{+\infty} e^u e^{-e^u}\, du \sim L(e^\mu). \tag{18.14}$$

Here $M(\xi)$ is introduced instead of $L(\xi)$ in order that we may have no trouble with finite singularities of $L(\xi)$.

To prove (18.14), let us reflect that

$$\begin{aligned} \int_0^\infty e^{(u-\mu)\alpha} e^{-e^{(u-\mu)}} M(e^u)\, du &= \int_0^{\mu-A} e^{(u-\mu)\alpha} e^{-e^{(u-\mu)}} M(e^u)\, du \\ &+ \int_{\mu-A}^{\mu+A} e^{(u-\mu)\alpha} e^{-e^{(u-\mu)}} M(e^u)\, du \\ &+ \int_{\mu+A}^{\infty} e^{(u-\mu)\alpha} e^{-e^{(u-\mu)}} M(e^u)\, du. \end{aligned} \tag{18.141}$$

Since $M(\xi)$ is asymptotic to $L(\xi)$, which is asymptotically increasing, the first integral is asymptotically less than

$$\begin{aligned} &L(e^\mu) \int_0^{\mu-A} e^{(u-\mu)\alpha} e^{-e^{(u-\mu)}}\, du \\ &\leqq L(e^\xi) \int_{-\infty}^{-A} e^u e^{-e^u}\, du \leqq L(e^\mu)\, e^{-A}. \end{aligned} \tag{18.142}$$

As to the second integral, it is asymptotically

$$L(e^\mu) \int_{-A}^{A} e^{\alpha u} e^{-e^u}\, du, \tag{18.143}$$

while the third one is ultimately less than

$$\begin{aligned} &e^{A^2} e^{-e^A} \int_{\mu+A}^{\infty} e^{-(u-\mu)} L(e^u)\, du \\ &\leqq e^{A^2} e^{-e^A} e^\mu \int_\mu^{+\infty} L(e^u)\, d(-e^{-u}) \\ &= e^{A^2} e^{-e^A} \left[L(e^\mu) + e^\mu \int_\mu^{+\infty} e^{-u} \frac{d}{du} (L(e^u))\, du \right] \\ &\leqq e^{A^2} e^{-e^A} \left[L(e^\mu) + C e^\mu \int_\mu^\infty e^{-u} u^{\alpha_1 - 1/2}\, du \right]. \end{aligned} \tag{18.144}$$

Now,

$$\begin{aligned}\int_\mu^\infty e^{-u}\, u^{\alpha_1-1/2}\, du &= -\int_\mu^\infty u^{\alpha_1-1/2}\, d(e^{-u}) \\ &= e^{-\mu}\,\mu^{\alpha_1-1/2} + \int_\mu^\infty (\alpha_1 - \tfrac{1}{2})\, e^{-u}\, u^{\alpha_1-3/2}\, du \\ &\leqq e^{-\mu}\left(\mu^{\alpha_1-1/2} + \int_\mu^\infty (\alpha_1-\tfrac{1}{2})\, u^{\alpha_1-3/2}\, du\right) \\ &\leqq 2e^{-\mu}\,\mu^{\alpha_1-1/2}.\end{aligned} \tag{18.145}$$

Thus by (18.144)

$$\int_{\mu+A}^\infty e^{(u-\mu)\alpha}\, e^{-e^{(u-\mu)}}\, M(e^u)\, du \leqq e^{A^2}\, e^{-e^A}\,[L(e^\mu) + o(L(e^\mu))]. \tag{18.146}$$

Combining this with (18.141), (18.142) and (18.143), we see that

$$\overline{\lim_{\mu\to\infty}} \left| \int_0^\infty e^{(u-\mu)\alpha}\, e^{-e^{(u-\mu)}}\, M(e^u)\, du - L(e^\mu)\int_{-\infty}^\infty e^u\, e^{-e^u}\, du \right| \Big/ L(e^\mu) \leqq \varphi(A) \tag{18.147}$$

where $\varphi(A)$ vanishes as A becomes infinite. Since A is arbitrarily large, (18.14) follows.

Consequently, by (18.14), if we put

$$d\psi(u) = \frac{1}{u}\, d\sum_0^{[u]} \frac{a_n}{n^{\alpha-1}} - A\, M(u), \tag{18.15}$$

(18.12) becomes

$$\int_0^\infty (w\xi)^\alpha\, e^{-w\xi}\, d\psi(w) \to 0. \tag{18.16}$$

On the other hand, we may write (18.11) in the form

$$\int_0^{1/\xi} (w\xi)^\alpha\, d\psi(w) \to 0. \tag{18.17}$$

Both integrals, (18.16) and (18.17), converge absolutely. We wish to make the transition from (18.16) to (18.17).

If $\alpha_1 \leqq 1$, $\psi(w)$ is "slowly decreasing" in the sense of Schmidt, we can apply Theorem XV, and it appears that $\psi(\lambda) \to 0$ as $\lambda \to \infty$. From this it follows that

$$\int_0^{1/\xi} (w\xi)^\alpha\, d\psi(w) = \psi\left(\frac{1}{\xi}\right) - \psi(0) - \alpha \int_0^{1/\xi} \xi^\alpha\, w^{\alpha-1}\, \psi(w)\, dw \to 0. \tag{18.18}$$

We now come to the more general case where $\alpha_1 > 1$. Our hypothesis becomes

$$e^{-\mu} \sum \frac{a_n}{n^{\alpha-1}(\log n)^{[\alpha_1]}}\, e^{(\log u-\mu)(\alpha-1)}\, e^{-e^{\log n-\mu}}\, \frac{\log n^{[\alpha_1]}}{\mu} \sim A\, M(e^\mu), \tag{18.19}$$

where

$$\Lambda(u) = (\log u)^{\alpha_1-[\alpha_1]} (\log\log u)^{\alpha_2} \dots (\log^{(n)} u)^{\alpha_n}. \tag{18.20}$$

The kernel of (18.19), in the sense of § 17, is

$$\begin{aligned} e^{(\alpha-1)(y-x)}\, e^{-e^{y-x}} \left\{\frac{y}{x}\right\}^{[\alpha_1]} &= e^{(\alpha-1)(y-x)}\, e^{-e^{y-1}} \left\{1-\frac{x-y}{x}\right\}^{[\alpha_1]} \\ &+ \left\{-[\alpha_1]\frac{x-y}{x} + \frac{[\alpha_1]\{[\alpha_1]-1\}}{2}\,\frac{(x-y)^2}{x^2} + \dots + \frac{(x-y)}{x^{[\alpha_1]}}\right\}. \end{aligned} \tag{18.21}$$

It hence appears that this kernel is of the form indicated in (17.04) where K_1 and K_1^* satisfy the appropriate conditions. Thus if we put $M^*(\xi) \sim \Lambda(\xi)$ at ∞ and $M^*(\xi)$ is continuous, and if

$$d\psi^*(u) = \frac{1}{u}\, d \sum_0^{[u]} \frac{a_n}{n^{\alpha-1}(\log n)^{[\alpha_1]}} - A M^*(u), \tag{18.22}$$

it follows just as in the case where $\alpha_1 \leqq 1$ that

$$\int_0^{1/\xi} (w\xi)^\alpha\, d\psi^*(w) \to 0. \tag{18.23}$$

This however is only another way of writing (18.11).

19. **The Tauberian theorem of Borel summation.**[52] The Borel sum of the series with partial sums s_m is

$$\lim_{x\to\infty} e^{-x} \sum_0^\infty \frac{s_n x^n}{n!} = s. \tag{19.01}$$

Let us put $n = [u^2]$ and

$$\sum_1^n s_m (\sqrt{m} - \sqrt{m-1}) = f(u) \tag{19.02}$$

for $u > 0$. Let $f(u) = 0$ for negative u. Let us also put $x = y^2$. Then (19.01) becomes

$$\lim_{y\to\infty} \int_0^\infty \frac{e^{-y^2} y^{2u^2}}{\Gamma(u^2+1)}\, \frac{df(u)}{u-\sqrt{u^2-1}} = s. \tag{19.03}$$

Now

$$\begin{aligned} \frac{e^{-y^2} y^{2u^2}}{\Gamma(u^2+1)(u-\sqrt{u^2-1})} &= \sqrt{\frac{2}{\pi}}\, e^{u^2-y^2} \left(\frac{y}{u}\right)^{2u^2} \left(1+O\left(\frac{1}{u^2}\right)\right) \\ &= \sqrt{\frac{2}{\pi}}\, e^{-2(u-y)^2}\, e^{3u^2-4uy+y^2} \left(\frac{y}{u}\right)^{2u^2} \left(1+O\left(\frac{1}{u^2}\right)\right). \end{aligned} \tag{19.04}$$

[52] For this section, cf. Schmidt (2); Vijayaraghavan (1).

Thus over the range $y - y^{1/6} \leqq u \leqq y + y^{1/6}$,

$$\frac{e^{-y^2} y^{2u^2}}{\Gamma(u^2+1)\,(u-\sqrt{u^2-1}\,)} - \sqrt{\frac{2}{\pi}}\, e^{-2(u-y)^2} = \sqrt{\frac{2}{\pi}}\, e^{-2(u-y)^2} O\left(\frac{1}{y^2}\right) \tag{19.05}$$

because

$$\begin{aligned} & e^{3u^2-4uy+y^2} \left(\frac{y}{u}\right)^{2u^2} \\ &= \exp\left\{3u^2 - 4uy + y^2 + 2u^2\left[\frac{y-u}{u} - \frac{(y-u)^2}{2u^2} + O\left(\frac{(y-u)^3}{u^3}\right)\right]\right\} \\ &= 1 + O\left(\frac{(y-u)^3}{u}\right) \\ &= 1 + O(y^{-1/2}). \end{aligned} \tag{19.06}$$

Moreover, it is always true that

$$\begin{aligned} \frac{e^{-y^2} y^{2u^2}}{\Gamma(u^2+1)\,(u-\sqrt{u^2-1}\,)} &= \sqrt{\frac{2}{\pi}}\, e^{u^2-y^2} \left(\frac{y}{u}\right)^{2u^2} \left(1 + O\left(\frac{1}{u^2}\right)\right) \\ &\leqq \sqrt{\frac{2}{\pi}}\, e^{u^2-y^2} e^{\left(\frac{y-u}{u}\right)2u^2} \left(1 + O\left(\frac{1}{u^2}\right)\right) \\ &= \sqrt{\frac{2}{\pi}}\, e^{-(u-y)^2} \left(1 + O\left(\frac{1}{u^2}\right)\right). \end{aligned} \tag{19.07}$$

Thus the conditions (17.05) and (17.06) are satisfied if we put

$$\frac{e^{-y^2} y^{2u^2}}{\Gamma(u^2+1)\,(u-\sqrt{u^2-1}\,)} = K_1(u, y) \tag{19.08}$$

and

$$\sqrt{\frac{2}{\pi}}\, e^{-2(u-y)^2} = K_1(u-y). \tag{19.09}$$

It follows at once that

$$\lim_{y\to\infty} \sqrt{\frac{2}{\pi}} \int_{-\infty}^{\infty} e^{-2(u-y)^2}\, df(u) = s \tag{19.10}$$

and that

$$\sqrt{\frac{2}{\pi}} \int_{-\infty}^{\infty} e^{-2(u-y)^2}\, df(u) \tag{19.101}$$

is bounded.

We now introduce again in its appropriate form Robert Schmidt's definition of a "langsam abfallende Folge". The sequence $\{s_m\}$ has this property if, whenever $q = q(p)$ $(p = 0, 1, \cdots)$ runs through such a sequence of indices that

$$q \geqq p \quad \text{and} \quad \frac{q-p}{\sqrt{p}} \to 0 \tag{19.11}$$

then

$$\lim_{p\to\infty}(s_q - s_p) \geqq 0. \tag{19.12}$$

Schmidt proves by elementary methods that if $s_0, s_1, \cdots$ is such a sequence, then

$$C(\Lambda) = \lim_{\substack{p\to\infty \\ p\leqq q\leqq p+\Lambda\sqrt{p}}} (s_q - s_p) \tag{19.13}$$

exists, and

$$\lim_{\Lambda\to 0} C(\Lambda) = 0. \tag{19.14}$$

Furthermore, he shows that there exists a constant K, such that

$$s_q - s_p \geqq -K(\sqrt{q} - \sqrt{p}) \tag{19.15}$$

for all $p = 0, 1, \cdots$ and all $q \geqq p + \sqrt{p}$. Thus he shows that a K exists such that

$$\frac{s_p}{\sqrt{p+1}} \geqq -K \qquad (p = 0, 1, \cdots). \tag{19.16}$$

Let the sequence s_m be "langsam abfallend" and let (19.01) hold. Still following Schmidt, we see that

$$\begin{aligned} e^{-x}\sum_0^\infty \frac{s_n x^n}{n!} &= \sum_0^{[x]} \frac{e^{-x}s_n x^n}{n!} + \sum_{[x]+1}^{[x+\sqrt{x}]} \frac{e^{-x}s_n x^n}{n!} + \sum_{[x+\sqrt{x}]+1}^{\infty} \frac{e^{-x}s_n x^n}{n!} \\ &\geqq -K\sum_0^\infty \frac{e^{-x}\sqrt{n+1}\, x^n}{n!} + s_{[x]} \sum_{[x]+1}^{[x+\sqrt{x}]} \frac{e^{-x}x^n}{n!} - C(1) \sum_{[x]+1}^{[x+\sqrt{x}]} \frac{e^{-x}x^n}{n!}. \end{aligned} \tag{19.17}$$

If we build up the appropriate $f(u)$ as in (19.02), and make the transformation which led from (19.01) to (19.03), we get:

$$e^{-x}\sum_0^\infty \frac{\sqrt{n+1}\, x^n}{n!} \leqq \int_0^\infty \frac{e^{-x}x^{u^2}}{\Gamma(u^2+1)} \frac{d(u^2)}{u - \sqrt{u^2-1}}. \tag{19.18}$$

Again, by (19.07)

$$e^{-x}\sum_0^\infty \frac{\sqrt{n+1}\, x^n}{n!} \leqq \int_{-\infty}^\infty e^{-(u-\sqrt{x})^2} O(u)\, du = O(\sqrt{x}). \tag{19.19}$$

Obviously

$$\sum_{[x]+1}^{[x+\sqrt{x}]} \frac{e^{-x}x^n}{n!} < 1. \tag{19.20}$$

On the other hand

$$\lim_{x\to\infty} \sum_{[x]+1}^{[x+\sqrt{x}]} \frac{e^{-x}x^n}{n!} > \lim_{x\to\infty} \frac{e^{-x}x^x}{e^{-x-\sqrt{x}}\,(x+\sqrt{x})^x\sqrt{2\pi(x+\sqrt{x})}}\,[\sqrt{x+1}]$$

$$(19.21)\qquad = \lim_{x\to\infty} \frac{e^{\sqrt{x}}}{\left(1+\frac{1}{\sqrt{x}}\right)^x\sqrt{2\pi}} = \left(\frac{e}{2\pi}\right)^{1/2}.$$

Thus (19.17) yields

$$(19.22)\qquad \overline{\lim_{x\to\infty}}\, x^{-1/2}\, s_{[x]} \leqq \overline{\lim_{x\to\infty}} \sqrt{\frac{2\pi}{e}}\,[s x^{-1/2} + KO(1) + C(1)\, x^{-1/2}] = O(1),$$

which we may simplify and write

$$(19.23)\qquad s_{[x]} = O(x^{1/2}).$$

Let us now introduce the function

$$(19.24)\qquad g_\alpha(x) = \frac{1}{\alpha^2}\left[\int_{x+\alpha}^{x+2\alpha} - \int_x^{x+\alpha}\right] f(y)\,dy.$$

By (19.02),

$$(19.25)\qquad g_\alpha(x) = \frac{1}{\alpha^2}\left[\int_{(x+\alpha)^2}^{(x+2\alpha)^2} - \int_{x^2}^{(x+\alpha)^2}\right] \frac{dM}{2\sqrt{M}} \sum_1^{[M]} s_m(\sqrt{m} - \sqrt{m-1}).$$

Now,

$$(19.26)\qquad \begin{aligned} &\int_0^\nu \frac{dM}{2\sqrt{M}} \sum_1^{[M]} s_m(\sqrt{m} - \sqrt{m-1}) \\ &\quad = \sum_1^{[\nu]} s_m(\sqrt{m} - \sqrt{m-1}) \int_m^\nu \frac{dM}{2\sqrt{M}} \\ &\quad = \sum_1^{[\nu]} s_m(\sqrt{m} - \sqrt{m-1})(\sqrt{\nu} - \sqrt{m}). \end{aligned}$$

Substituting this in (19.25), and writing S for an average of the s_n's for $x^2 < n < (x+2\alpha)^2 + 1$, we have

$$\begin{aligned} g_\alpha(x) &= \frac{1}{\alpha^2} \sum_1^{[(x+2\alpha)^2]} s_m(\sqrt{m} - \sqrt{m-1})(x+2\alpha-\sqrt{m}) \\ &\quad - \frac{2}{\alpha^2} \sum_1^{[(x+\alpha)^2]} s_m(\sqrt{m} - \sqrt{m-1})(x+\alpha-\sqrt{m}) \\ &\quad + \frac{1}{\alpha^2} \sum_1^{[x^2]} s_m(\sqrt{m} - \sqrt{m-1})(x-\sqrt{m}) \\ &= \frac{1}{\alpha^2}\left\{ \sum_{[x^2]+1}^{[(x+\alpha)^2]} s_m(\sqrt{m} - \sqrt{m-1})(\sqrt{m} - x) \right. \\ &\quad \left. + \sum_{[(x+\alpha)^2]+1}^{[(x+2\alpha)^2]} s_m(\sqrt{m} - \sqrt{m-1})(x+2\alpha-\sqrt{m}) \right\} \end{aligned}$$

$$
\begin{aligned}
(19.27) \qquad &= \frac{S}{\alpha^2}\left\{\sum_{[x^2]+1}^{[(x+\alpha)^2]}(\sqrt{m}-\sqrt{m-1})(\sqrt{m}-x)\right. \\
&\quad \left.+\sum_{[(x+\alpha)^2]+1}^{[(x+2\alpha)^2]}(\sqrt{m}-\sqrt{m-1})(x+2\alpha-\sqrt{m})\right\} \\
&= \frac{S}{\alpha^2}\left[\int_{x+\alpha}^{x+2\alpha}-\int_{x}^{x+\alpha}\right]\sqrt{[x^2]}\,dx \\
&= S\left\{1+\frac{1}{\alpha^2}\left[\int_{x+\alpha}^{x+2\alpha}-\int_{x}^{x+\alpha}\right](\sqrt{[x^2]}-x)\,dx\right\} \\
&= S\left\{1+O\left(\frac{1}{x^2}\right)+\frac{1}{\alpha^2}\left[\sum_{[x^2]}^{[(x+\alpha)^2]}-\sum_{[(x+\alpha)^2]}^{[(x+2\alpha)^2]}\right]\frac{(\sqrt{m+1}-\sqrt{m})^2}{2}\right\} \\
&= S\left(1+O\left(\frac{1}{x^2}\right)\right).
\end{aligned}
$$

It is indeed easy to show that as x increases, the weight of the later s_n's averaged in S increases at the expense of the weight of the earlier ones.

We may show that the condition that $\{s_m\}$ is "langsam abfallend" may be put in the form that if

$$(19.28) \qquad v \geqq u, \quad v-u \to 0$$

then

$$(19.29) \qquad \lim_{u\to\infty}(s_{[v^2]}-s_{[u^2]}) \equiv 0.$$

Hence by (19.27) and (19.23), we may readily show that

$$(19.30) \qquad \lim_{u\to\infty}(g_\alpha(v)-g_\alpha(u)) \geqq 0.$$

Again remembering (19.23), which makes all the integrals in question absolutely convergent, we have

$$
\begin{aligned}
&\lim_{y\to\infty}\sqrt{\frac{2}{\pi}}\int_{-\infty}^{\infty}e^{-2(u-y)^2}g_\alpha(u)\,du \\
&= \lim_{y\to\infty}\sqrt{\frac{2}{\pi}}\int_{-\infty}^{\infty}e^{-2(u-y)^2}\,du\,\frac{1}{\alpha}\int_{u}^{u+\alpha}\frac{dv}{\alpha}\int_{v}^{v+\alpha}df(w) \qquad (19.31) \\
&= \lim_{y\to\infty}\frac{1}{\alpha^2}\sqrt{\frac{2}{\pi}}\int_{y}^{y+\alpha}dv\int_{v}^{v+\alpha}dw\int_{-\infty}^{\infty}e^{-2(u-w)^2}\,df(u) \\
&= \lim_{y\to\infty}\sqrt{\frac{2}{\pi}}\int_{-\infty}^{\infty}e^{-2(u-y)^2}\,df(u) = s.
\end{aligned}
$$

An integration by parts yields:

$$(19.32) \qquad s = \lim_{y\to\infty}\sqrt{\frac{2}{\pi}}\int_{-\infty}^{\infty}dg_\alpha(v)\int_{v}^{\infty}e^{-2(u-y)^2}\,du.$$

Similar methods establish the boundedness of

$$\sqrt{\frac{2}{\pi}} \int_{-\infty}^{\infty} d\, g_\alpha(v) \int_v^{\infty} e^{-2(u-v)^2}\, du. \tag{19.33}$$

We may then apply Theorem XIV, taking

$$M(x) = \sqrt{\frac{2}{\pi}} \int_x^{\infty} e^{-2u^2}\, du. \tag{19.34}$$

In this case

$$\varphi(z) = \sqrt{\frac{1}{2\pi}}\, e^{z^2/8}. \tag{19.35}$$

Hence

$$\lim_{x \to \infty} g_\alpha(x) = s. \tag{19.36}$$

Let us now return to (19.27), making use of the existence of (19.13) and of (19.14). Then if x is sufficiently large,

$$S - s_{[x^2]} \geqq C(5\alpha) \tag{19.37}$$

and

$$s_{[(x+2\alpha)^2]} - S \geqq C(5\alpha). \tag{19.38}$$

It then follows from (19.27) and (19.36) that

$$\overline{\lim_{n \to \infty}}\, s_n \leqq s - C(5\alpha) \tag{19.39}$$

and

$$\underline{\lim_{n \to \infty}}\, s_n \geqq s + C(5\alpha). \tag{19.40}$$

Combining these, and remembering that $C(\mathcal{A}) \leqq 0$, we see that the following theorem holds:

THEOREM XXI. *Let* (19.01) *hold, and let* s_m *be "langsam abfallend". Then*

$$\lim_{n \to \infty} s_n = s. \tag{19.41}$$

CHAPTER VII.

A QUASI-TAUBERIAN THEOREM.

20. **The quasi-Tauberian theorem.** Up to the present point, all the Tauberian theorems we have discussed have involved some auxiliary condition of boundedness or positiveness or slow decrease. In the present chapter, we shall discuss a theorem without any such auxiliary condition as to the function averaged. The type of theorem is so fundamentally

different from that already discussed that the nomenclature, "Tauberian", seems to the author unfortunate. The theorems now to be discussed are in essence much closer to Abel's theorem than to Tauber's theorem.

By an integration by parts, we establish the following:

LEMMA XXIIa. *Let*

$$\int_B^C |F(x)|\,|df(x)| \tag{20.01}$$

exist for every $C>B$, *and let*

$$\int_B^\infty F(x)\,df(x) \tag{20.02}$$

exist as the limit of $\int_B^C F(x)\,df(x)$ *as* $C\to\infty$. *Let*

$$\int_B^C |G(x)|\,|df(x)| \tag{20.03}$$

exist for each $C>B$, *let* $G(x)$ *and* $F(x)$ *be continuous, and let*

$$\frac{G(x)}{F(x)}\to L \text{ as } x\to\infty; \qquad \int_D^B \left| d\,\frac{G(x)}{F(x)}\right| <\infty. \tag{20.04}$$

Then

$$\int_B^\infty G(x)\,df(x) \tag{20.05}$$

will exist, and

$$\begin{aligned}&\int_B^\infty G(x)\,df(x)\\ &= L\int_B^\infty F(x)\,df(x) - \int_B^\infty d\left(\frac{G(x)}{F(x)}\right)\int_B^x F(\xi)\,df(\xi).\end{aligned} \tag{20.06}$$

Now let us suppose that $f(x)$ is of limited total variation over every finite interval, that $K_1(x)$ is bounded and continuous, that

$$\int_{-\infty}^\infty |d(K_1(x)\,e^{-\lambda x})| < \text{const.}, \tag{20.061}$$

and that

$$\int_D^\infty K_1(-x)\,df(x) \tag{20.07}$$

exists in the sense of (20.02). Let us suppose further that as $x\to-\infty$

$$K_1(x)\sim A_1 e^{\lambda x}, \qquad (\lambda>0) \tag{20.08}$$

where $A_1\neq 0$. Then

$$\int_{-\infty}^{-D} \left| d \frac{e^{\lambda x}}{K_1(x)} \right| < \text{const.}, \tag{20.081}$$

if D is sufficiently large that for $x \leqq -D$,

$$|e^{-\lambda x} K_1(x)| \geqq \text{const.} > 0. \tag{20.082}$$

As a particular case of (20.06), if B is large enough,

$$\begin{aligned} &\int_B^{\infty} e^{-\lambda x} \, d f(x) \\ &= \frac{1}{A_1} \int_B^{\infty} K_1(-x) \, d f(x) - \int_B^{\infty} d \left(\frac{e^{-\lambda x}}{K_1(-x)} \right) \int_B^{x} K_1(-\xi) \, d f(\xi) \end{aligned} \tag{20.083}$$

and hence

$$\int_B^{\infty} e^{-\lambda x} \, d f(x) \tag{20.084}$$

is bounded. A further application of (20.06) yields us

$$\begin{aligned} &\int_B^{\infty} K_1(y-x) \, d f(x) \\ &= A_1 e^{\lambda y} \int_B^{\infty} e^{-\lambda x} \, d f(x) - \int_B^{\infty} d(K_1(y-x) \, e^{\lambda x}) \int_B^{x} e^{-\lambda \xi} \, d f(\xi). \end{aligned} \tag{20.085}$$

From this we may conclude that

$$\int_0^{\infty} K_1(y-x) \, d f(x) = \begin{cases} O(e^{\lambda y}) & \text{as } y \to \infty, \\ O(1) & \text{as } y \to -\infty \end{cases} \tag{20.09}$$

and that

$$\lim_{B \to \infty} e^{-\lambda y} \int_B^{\infty} K_1(y-x) \, d f(x) = 0 \tag{20.10}$$

uniformly in y.

Now let $R(z)$ be a function for which

$$\int_{-\infty}^{\infty} e^{-\lambda z} \, |dR(z)| \tag{20.11}$$

and

$$\int_{-\infty}^{\infty} |dR(z)| \tag{20.12}$$

are finite. It follows from (20.10) that

$$\begin{aligned} &\int_{-\infty}^{\infty} dR(y-z) \int_0^{\infty} K_1(z-x) \, d f(x) \\ &= \lim_{B \to \infty} \int_{-\infty}^{\infty} dR(y-z) \int_0^{B} K_1(z-x) \, d f(x) \\ &= \lim_{B \to \infty} \int_0^{B} d f(x) \int_{-\infty}^{\infty} K_1(z-x) \, d R(y-z) \\ &= \int_0^{\infty} d f(x) \int_{-\infty}^{\infty} K_1(y-x-z) \, d R(z). \end{aligned} \tag{20.13}$$

Thus if

$$\lim_{y\to\infty}\int_0^\infty K_1(y-x)\,df(x) = A\int_{-\infty}^\infty K_1(x)\,dx \tag{20.14}$$

then

$$\int_0^\infty K_1(y-x)\,df(x) \tag{20.15}$$

will be bounded, by (20.09), and it follows that

$$\begin{aligned}
&\lim_{y\to\infty}\int_0^\infty df(x)\int_{-\infty}^\infty K_1(y-x-z)\,dR(z)\\
&\quad=\lim_{y\to\infty}\int_{-\infty}^\infty dR(y-z)\int_{-\infty}^\infty K_1(z-x)\,df(x)\\
&\quad=\int_{-\infty}^\infty dR(-z)\lim_{y\to\infty}\int_0^\infty K_1(z+y-x)\,df(x) \qquad (20.16)\\
&\quad=A\int_{-\infty}^\infty K_1(x)\,dx\int_{-\infty}^\infty dR(-z)\\
&\quad=A\int_{-\infty}^\infty dx\int_{-\infty}^\infty K_1(x-z)\,dR(z).
\end{aligned}$$

Hence if $K_1(x)$ *is bounded and continuous, if* $f(x)$ *is of limited total variation over every finite interval, if* (20.14), (20.061) *and* (20.08) *hold, if* (20.11) *and* (20.12) *are finite, and if*

$$K_2(x) = \int_{-\infty}^\infty K_1(x-z)\,dR(z), \tag{20.17}$$

it follows that

$$\lim_{y\to\infty}\int_0^\infty K_2(y-x)\,df(x) = A\int_{-\infty}^\infty K_2(x)\,dx. \tag{20.18}$$

This is a sufficiently important theorem to dignify by a number; we shall call it Theorem XXII.

A closely related proposition is:

THEOREM XXIII. *In the hypothesis of Theorem* XXII, *in case*

$$K_1(x) = 0, \qquad [x>0], \tag{20.19}$$

we may replace the assumption of the finiteness of (20.11) *and* (20.12) *by that of the finiteness of*

$$\int_{-\infty}^\infty |dR(z)|. \tag{20.20}$$

The conclusion remains valid.

To prove this, let us reflect that if (20.08) holds as $x \to \infty$, since

$$\int_0^\infty K_1(y-x)\,df(x) = \int_y^\infty K_1(y-x)\,df(x), \tag{20.21}$$

it will follow from (20.06) that

$$\begin{aligned}\int_y^\infty e^{-\lambda x}\,df(x) &= \frac{e^{-\lambda y}}{A_1}\int_y^\infty K_1(y-x)\,df(x)\\ &\quad -\int_y^\infty d\left(\frac{e^{-\lambda x}}{K_1(y-x)}\right)\int_y^x K_1(y-\xi)\,df(\xi)\\ &= e^{-\lambda y}\,O(1).\end{aligned} \tag{20.22}$$

By a precisely similar argument,

$$\begin{aligned}&\int_B^\infty K_1(y-x)\,df(x)\\ &= A_1 e^{\lambda y}\int_B^\infty e^{-\lambda x}\,df(x) - \int_B^\infty d_x\left(e^{\lambda x}K_1(y-x)\right)\int_B^\infty e^{-\lambda\xi}\,df(\xi)\\ &= A_1 e^{\lambda(y-B)}O(1) - \int_B^\infty d_x\left(e^{\lambda x}K_1(y-x)\right)\left(e^{-\lambda B}O(1)+e^{-\lambda x}O(1)\right)\\ &= O(e^{\lambda(y-B)}) \quad as \quad B\to\infty\end{aligned} \tag{20.23}$$

if $B>y$. Thus by (20.21), $\int_B^\infty K_1(y-x)\,df(x)$ is less than a function which is bounded and decreases monotonely to 0 as B becomes infinite. Consequently, by a theorem of Daniell on the Stieltjes integral,[53] in combination with the finiteness of (20.21),

$$\lim_{B\to\infty}\int_{-\infty}^\infty dR(y-z)\int_B^\infty K_1(z-x)\,df(x) = 0. \tag{20.24}$$

Thus the inversions of integration in (20.13) are again permissible.

Formally and heuristically, (20.17) is equivalent to

$$\frac{\int_{-\infty}^\infty K_2(x)e^{ux}\,dx}{\int_{-\infty}^\infty K_1(x)e^{ux}\,dx} = \int_{-\infty}^\infty e^{ux}\,dR(x). \tag{20.25}$$

If K_1 and K_2 are both $O(e^{-(u+\varepsilon)x})$ at $+\infty$ and $O(e^{+\varepsilon x})$ at $-\infty$,

$$k_2(u) = \int_{-\infty}^\infty K_2(x)\,e^{ux}\,dx \tag{20.26}$$

and

$$k_1(u) = \int_{-\infty}^\infty K_1(x)\,e^{ux}\,dx \tag{20.27}$$

[53] Daniell (1).

are both analytic over $-\varepsilon < \Re(u) < \varepsilon + \mu$. Formally,

$$R(x) = \frac{1}{2\pi i} \int_0^x d\xi \int_{-\infty}^{\infty} \frac{k_2(u)}{k_1(u)} e^{-u\xi}\, du, \tag{20.28}$$

the second integral being taken along any ordinate in the strip $-\varepsilon < \Re(u) < \varepsilon + \mu$.

Now let $K_1(x)$ belong to L_2. Let $k_2(u)/k_1(u)$ be analytic over $-\varepsilon \leqq \Re(u) \leqq \mu + \varepsilon$, and let it be quadratically summable over every ordinate in that strip. Then

$$e^{-\varepsilon x}\, \text{l. i. m.}_{A\to\infty} \int_{-A}^{A} \frac{k_2(iu)}{k_1(iu)} e^{-iux}\, du \tag{20.29}$$

and

$$e^{(\mu+\varepsilon)x}\, \text{l. i. m.}_{A\to\infty} \int_{-A}^{A} \frac{k_2(iu)}{k_1(iu)} e^{-iux}\, du \tag{20.30}$$

are quadratically summable. As a consequence,

$$\text{l. i. m.}_{A\to\infty} \int_{-A}^{A} \frac{k_2(iu)}{k_1(iu)} e^{-iux}\, du \tag{20.31}$$

and

$$e^{\mu x}\, \text{l. i. m.}_{A\to\infty} \int_{-A}^{A} \frac{k_2(iu)}{k_1(iu)} e^{-iux}\, du \tag{20.32}$$

are absolutely summable. If we assume R to be defined as in (20.28) and $\mu = \lambda$, we obtain:

THEOREM XXII′. *Let $K_1(x)$ be bounded and continuous. Let $f(x)$ be of limited total variation over every finite interval. Let* (20.14), (20.061) *and* (20.08) *hold. Let $k_2(u)$ and $k_1(u)$ be defined as in* (20.26) *and* (20.27), *respectively. Let $K_1(x)$ belong to L_2. Let $k_2(u)/k_1(u)$ be analytic over $-\varepsilon \leqq \Re(u) \leqq \lambda + \varepsilon$, and let it belong to L_2 over every ordinate in that strip. Then* (20.18) *follows.*

Similarly, we have:

THEOREM XXIII′. *In the hypothesis of Theorem* XXII′, *in case $K_1(x)$ vanishes for positive arguments, we may replace the strip $-\varepsilon \leqq \Re(u) \leqq \lambda + \varepsilon$ by the narrower strip $-\varepsilon \leqq \Re(u) \leqq \varepsilon$.*

21. **Applications of the quasi-Tauberian theorem.** If

$$K^{(m)}(x) = \begin{cases} (m+1)(1-e^x)^m e^x; & [x<0] \\ 0 & [x>0] \end{cases} \tag{21.01}$$

and

$${}^{(m)}K(x) = \frac{2e^x}{\pi} \int_0^1 (1-z)^m \cos(z e^x)\, dz; \qquad [m>0] \tag{21.02}$$

then (20.08) is clearly satisfied for $\lambda = 1$, and

$$K_1(x) = K^{(m)}(x) \tag{21.03}$$

or

$$K_1(x) = {}^{(m)}K(x). \tag{21.04}$$

If we put

$$\tau_m(u) = \int_{-\infty}^{\infty} K^{(m)}(x)\, e^{iux}\, dx = \frac{\Gamma(m+2)\,\Gamma(u+1)}{\Gamma(u+m+2)}, \tag{21.05}$$

and

$$\sigma_m(u) = \int_{-\infty}^{\infty} {}^{(m)}K(x)\, e^{iux}\, dx = \frac{\Gamma(m+1)}{\Gamma(m+1-u)\cos\dfrac{\pi u}{2}}, \tag{21.06}$$

then as $|\Im(u)| \to \infty$

$$\left|\frac{\sigma_m(u)}{\tau_n(u)}\right| \sim \frac{2\,\Gamma(m+1)}{\sqrt{2\pi}\,\Gamma(n+2)\,e^{\Re(u)-m}}\,|\Im(u)|^{\Re(u)-m+n+1/2}. \tag{21.07}$$

Thus if $n > m$, then

$$\lim_{y\to\infty} \int_0^{\infty} {}^{(m)}K(y-x)\, df(x) = A\int_0^{\infty} {}^{(m)}K(x)\, dx \tag{21.08}$$

will imply

$$\lim_{y\to\infty} \int_0^{\infty} K^{(n)}(y-x)\, df(x) = A\int_{-\infty}^{\infty} K^{(n)}(x)\, dx, \tag{21.09}$$

while if $m > n+1$, (21.09) will imply (21.08).

With a little manipulation, already indicated by the authors in question, this result is seen to be equivalent to an important theorem of Hardy and Littlewood,[54] which gives the necessary and sufficient condition for the summability of Fourier series and integrals by Cesàro sums of some order. In the integral form, the theorem reads as follows: *Let $f(x)$ be a measurable function defined over $(-\infty, \infty)$, and zero outside $(-A, A)$. Let*

$$\varphi(y) = \tfrac{1}{2}\{f(x+y)+f(x-y)-2s\}. \tag{21.10}$$

Then if we write B_m for the proposition

$$\lim_{\lambda\to\infty} \lambda \lim_{\varepsilon\to 0} \int_{\varepsilon}^{1/\lambda} \varphi(y)\,(1-\lambda y)^{m-1}\, dy = 0, \tag{21.11}$$

and C_m for the proposition

$$\lim_{\lambda\to\infty} \lambda \lim_{\varepsilon\to 0} \int_{\varepsilon}^{\infty} \varphi(y)\, dy \int_0^1 (1-z)^m \cos\lambda yz\, dz = 0, \tag{21.12}$$

B_m implies $C_{m+\varepsilon}$ for $m \geqq 1$, while C_m implies $B_{m+1+\varepsilon}$ for $m > 0$.

[54] Hardy and Littlewood (12).

While the general method of the present section has been developed in an earlier paper of the author,[55] his final results were not stated correctly. The correct result is due to Bosanquet[56] and Paley, who have shown it to be a "best possible" result in both directions. Their theorem also applies to $m > -1$, and is otherwise somewhat more general.

CHAPTER VIII.

TAUBERIAN THEOREMS AND SPECTRA.

22. A further type of asymptotic behavior. The O and o symbols do not exhaust the possible terms in which we may describe the behavior of a function at infinity. A proposition which may be regarded as in some wise a generalization of

$$\lim_{x \to \infty} f(x) = A \tag{22.01}$$

is

$$\lim_{B \to \infty} \frac{1}{B} \int_0^B |f(x) - A|^2 \, dx = 0. \tag{22.02}$$

If $f(x)$ is bounded and measurable, (22.01) clearly implies (22.02), while (22.02) does not imply (22.01). Proposition (22.02) has a certain analogy to the different types of "strong convergence" to a limit which a function may exhibit: namely

$$\int_0^\infty |f(x) - A| \, dx \text{ converges} \tag{22.03}$$

and

$$\int_0^\infty |f(x) - A|^2 dx \text{ converges.} \tag{22.04}$$

The series analogues of the latter

$$\sum_1^\infty |s_n - A| \text{ converges} \tag{22.05}$$

and

$$\sum_1^\infty |s_n - A|^2 \text{ converges} \tag{22.06}$$

imply the ordinary convergence of s_n to A, but are not implied by it. On the other hand, neither (22.03) nor (22.04) is implied by ordinary convergence, although (22.04) implies (22.02). In contrast with propositions

[55] Wiener (5).

[56] Bosanquet (1), (2); Paley (1).

(22.03)—(22.06), which represent various types of "strong convergence", we shall express (22.02) in the usual language by saying that $f(x)$ is *strongly summable*[57] to A as $x \to \infty$. We shall say that A is a *sublimit* of $f(x)$, and shall write

$$\operatorname*{slm}_{x\to\infty} f(x) = A. \tag{22.07}$$

The sublimit of a function $f(x)$ differs markedly from the ordinary limit in that neither its existence nor its value are invariant if we replace x by a monotone function of x becoming infinite with x. On the other hand, the sublimit of $f(x)$ has a relation to the harmonic analysis of $f(x)$ far closer than does the ordinary limit.

Closely related to the notion of sublimit is that of subboundedness, which bears to the ordinary notion of boundedness much the same relation which that of sublimit does to the ordinary notion of limit. A function $f(x)$ is said to be *subbounded* if

$$\frac{1}{B}\int_0^B |f(x)|^2\,dx \tag{22.08}$$

is bounded. This juxtaposition of a notion of limit and a notion of boundedness suggests a generalized form of Tauberian theorem. To be specific, let us ask what conditions beyond the subboundedness of $f(x)$ are sufficient to make

$$\operatorname*{slm}_{x\to\infty} \int_0^\infty K_1(x-\xi) f(\xi)\,d\xi = A\int_{-\infty}^\infty K_1(\xi)\,d\xi \tag{22.09}$$

imply

$$\operatorname*{slm}_{x\to\infty} \int_0^\infty K_2(x-\xi) f(\xi)\,d\xi = A\int_{-\infty}^\infty K_2(\xi)\,d\xi. \tag{22.10}$$

This problem belongs to the range of ideas treated by the author in his work on generalized harmonic analysis. It gives a clearer picture of the real significance of Tauberian theorems. On the other hand, it does not at present offer an alternative approach to them, since Tauberian theorems of the type already discussed in this paper play an essential role in the establishment of a theory of generalized harmonic analysis. Results involving these theorems will be applied in this section.

To return to (22.09) and (22.10), no generality is lost by taking $A = 0$, as this simply amounts to replacing $f(x)$ by $f(x) - A$. Let us then take $A = 0$, and let us put $f(x) = 0$ for negative arguments, which

[57] Hardy and Littlewood (15).

is clearly permissible. We then wish to find a set of conditions which in conjunction with the boundedness for large B of

$$\frac{1}{2B}\int_{-B}^{B}|f(x)|^2\,dx \tag{22.11}$$

(or what is the same, with the fact that $\frac{f(x)}{1+|x|}$ belongs to L_2) and the proposition

$$\lim_{B\to\infty}\frac{1}{2B}\int_{-B}^{B}\left|\int_{-\infty}^{\infty}K_1(x-\xi)f(\xi)\,d\xi\right|^2dx = 0 \tag{22.12}$$

are sufficient to imply

$$\lim_{B\to\infty}\frac{1}{2B}\int_{-B}^{B}\left|\int_{-\infty}^{\infty}K_2(x-\xi)f(\xi)\,d\xi\right|^2d\xi = 0. \tag{22.13}$$

We shall prove the following:

THEOREM XXIV. *Let $f(x)$ be measurable, and let* (22.11) *be bounded. Let* (22.12) *be valid. Let K_1 and K_2 be measurable, and let*

$$\int_{-\infty}^{\infty}(1+|x|)^{3+\varepsilon}\,|K_1(x)|^2\,dx<\infty \quad \left(\text{and hence}\int_{-\infty}^{\infty}(1+|x|)\,|K_1(x)|\,dx<\infty\right), \tag{22.14}$$

and

$$\int_{-\infty}^{\infty}(1+|x|)^{3+\varepsilon}\,|K_2(x)|^2\,dx<\infty \quad \left(\text{and hence}\int_{-\infty}^{\infty}(1+|x|)\,|K_2(x)|\,dx<\infty\right). \tag{22.15}$$

Furthermore, let

$$\int_{-\infty}^{\infty}K_1(x)\,e^{iux}\,dx \neq 0 \tag{22.16}$$

for all real u. Then (22.13) *is true.*

For this theorem we shall need the following:

LEMMA XXIVa. *Let $f(x, u, \lambda)$ belong to L_2 as a function of u and be measurable in (x, u). Let $\int_{-\infty}^{\infty}|f(x, u, \lambda)|^2\,du$ be bounded, and let*

$$\int_{-\infty}^{\infty}|f(x, u, \lambda)-f(x, u)|^2\,du\to 0 \tag{22.17}$$

uniformly over any finite range of x as $\lambda\to\infty$. Let K be measurable, and let

(22.18) $$\int_{-\infty}^{\infty} |K(x)|\,dx$$

be finite. Let

(22.19) $$\int_{-\infty}^{\infty} K(x)\,f(x,u,\lambda)\,dx$$

exist for all λ. *Then*

$$(22.20)\quad \int_{-\infty}^{\infty}\left|\int_{-\infty}^{\infty} K(x)f(x,u,\lambda)\,dx-\int_{-\infty}^{\infty} K(x)f(x,u)\,dx\right|^2 du\to 0 \text{ as } \lambda\to\infty.$$

To prove this lemma, let us reflect that

$$\begin{aligned}
&\left[\int_A^B |K(x)|\,dx\right]^2 \max_x \int_{-\infty}^{\infty} |f(x,u,\lambda)-f(x,u)|^2\,du\\
&\quad\geqq \int_A^B |K(x)|\,dx\int_A^B |K(y)|\,dy\int_{-\infty}^{\infty} du\,|f(x,u,\lambda)-f(x,u)|\\
&\qquad\qquad\times|f(y,u,\lambda)-f(y,u)|\\
&\quad= \int_{-\infty}^{\infty} du\int_A^B |K(x)|\,dx\int_A^B |K(y)|\,dy\,|f(x,u,\lambda)-f(x,u)|\\
&\qquad\qquad\times|f(y,u,\lambda)-f(y,u)|\\
&\quad\geqq \int_{-\infty}^{\infty} du\left|\int_A^B K(x)\,[f(x,u,\lambda)-f(x,u)]\,dx\right|^2.
\end{aligned}\tag{22.21}$$

By taking A large and B infinite, or $-B$ large and $-A$ infinite, we may make the left-hand member uniformly small. Over any finite range (A, B) the left-hand member tends to 0 as $\lambda\to\infty$. By combining these facts the lemma follows. It will be noted that it involves the existence almost everywhere of

(22.22) $$\int_{-\infty}^{\infty} K(x)\,f(x,u)\,dx.$$

We may now return to the proof of our main theorem. In a previous paper[58], the author has shown that

$$(22.23)\quad s(u)=\frac{1}{2\pi}\int_{-1}^{1} f(x)\frac{e^{iux}-1}{ix}\,dx+\underset{A\to\infty}{\text{l.i.m.}}\frac{1}{\pi}\left[\int_1^A+\int_A^{-1}\right]f(x)\frac{e^{iux}}{ix}\,dx$$

exists. Moreover,

$$\begin{aligned}
&\int_{-\infty}^{\infty} |K_1(x-\xi)f(\xi)|\,d\xi \leqq \int_{-\infty}^{\infty} |K_1(\xi)|\,|f(x-\xi)|\,d\xi\\
&\quad\leqq \left[\int_{-\infty}^{\infty}(1+|\xi|)^2\,|K_1(\xi)|^2\,d\xi\int_{-\infty}^{\infty}\frac{|f(x-\xi)|^2}{(1+|\xi|)^2}\,d\xi\right]^{1/2}\\
&\quad<\infty,
\end{aligned}\tag{22.24}$$

[58] Wiener (7).

so that

$$\frac{1}{2B}\int_{-B}^{B}\left|\int_{-\infty}^{\infty}K_1(x-\xi)f(\xi)\,d\xi\right|^2 dx$$

$$\leqq \frac{1}{2B}\int_{-B}^{B}dx\int_{-\infty}^{\infty}|K_1(\xi)|\,|f(x-\xi)|\,d\xi\int_{-\infty}^{\infty}|K_1(\eta)|\,|f(x-\eta)|\,d\eta$$

$$= \frac{1}{2B}\int_{-\infty}^{\infty}|K_1(\xi)|\,d\xi\int_{-\infty}^{\infty}|K_1(\eta)|\,d\eta\int_{-B}^{B}|f(x-\xi)|\,|f(x-\eta)|\,dx \tag{22.25}$$

$$\leqq \frac{1}{2B}\int_{-\infty}^{\infty}|K_1(\xi)|\,d\xi\int_{-\infty}^{\infty}|K_1(\eta)|\,d\eta\int_{-B-|\xi|-|\eta|}^{B+|\xi|+|\eta|}|f(x)|^2\,dx$$

$$\leqq \frac{\text{const.}}{B}\int_{-\infty}^{\infty}|K_1(\xi)|\,d\xi\int_{-\infty}^{\infty}|K_1(\eta)|\,d\eta\,[B+|\xi|+|\eta|]$$

$$< \text{const.}$$

Thus the function

$$f_1(x) = \int_{-\infty}^{\infty}K_1(x-\xi)f(\xi)\,d\xi \tag{22.26}$$

and the analogous function

$$f_2(x) = \int_{-\infty}^{\infty}K_2(x-\xi)f(\xi)\,d\xi \tag{22.27}$$

define functions $s_1(u)$ and $s_2(u)$, corresponding to them in the same way in which $s(u)$ corresponds to $f(x)$.

In his previous paper,[58] the author has shown that

$$s(u+\varepsilon)-s(u-\varepsilon) = \frac{1}{\pi}\operatorname*{l.i.m.}_{A\to\infty}\int_{-A}^{A}f(x)\frac{\sin\varepsilon x}{x}e^{iux}\,dx. \tag{22.28}$$

Similarly

$$s_1(u+\varepsilon)-s_1(u-\varepsilon)$$

$$= \frac{1}{\pi}\operatorname*{l.i.m.}_{A\to\infty}\int_{-A}^{A}\frac{\sin\varepsilon x}{x}e^{iux}\,dx\int_{-\infty}^{\infty}K_1(\xi)f(x-\xi)\,d\xi. \tag{22.29}$$

There is no trouble in modifying (22.28) into

$$s(u+\varepsilon)-s(u-\varepsilon) = \frac{1}{\pi}\operatorname*{l.i.m.}_{A\to\infty}\int_{-A-\xi}^{A-\xi}f(x)\frac{\sin\varepsilon x}{x}e^{iux}\,dx. \tag{22.30}$$

Thus

$$s_1(u+\varepsilon)-s_1(u-\varepsilon)-(s(u+\varepsilon)-s(u-\varepsilon))\int_{-\infty}^{\infty}K_1(\xi)\,e^{iu\xi}\,d\xi$$

$$= \frac{1}{\pi}\left[\operatorname*{l.i.m.}_{A\to\infty}\int_{-\infty}^{\infty}K_1(\xi)\,d\xi\int_{-A}^{A}f(x-\xi)\frac{\sin\varepsilon x}{x}e^{iux}\,dx\right.$$

$$\left.-\int_{-\infty}^{\infty}K_1(\xi)\,d\xi\operatorname*{l.i.m.}_{A\to\infty}\int_{-A}^{A}f(x-\xi)\frac{\sin\varepsilon(x-\xi)}{x-\xi}e^{iux}\,dx\right] \tag{22.31}$$

$$= \frac{1}{\pi}\operatorname*{l.i.m.}_{A\to\infty}\int_{-\infty}^{\infty}K_1(\xi)\,d\xi\int_{-A}^{A}f(x-\xi)\left[\frac{\sin\varepsilon x}{x}-\frac{\sin\varepsilon(x-\xi)}{x-\xi}\right]e^{iux}\,dx.$$

In the proof of this we have used our last lemma.

Clearly

$$\left|\frac{\sin \varepsilon x}{\varepsilon x} - \frac{\sin \varepsilon(x-\xi)}{\varepsilon(x-\xi)}\right| \leqq 2. \tag{22.32}$$

Again,

$$\begin{aligned}\left|\frac{\sin \varepsilon x}{\varepsilon x} - \frac{\sin \varepsilon(x-\xi)}{\varepsilon(x-\xi)}\right| &= \left|\int_{x-\xi}^{x} \frac{d}{dw}\left(\frac{\sin \varepsilon w}{\varepsilon w}\right) dw\right| \\ &= \left|\int_{x-\xi}^{x} \frac{\varepsilon^2 w \cos \varepsilon w - \varepsilon \sin \varepsilon w}{\varepsilon^2 w^2}\, dw\right| \\ &\leqq \varepsilon|\xi| \max_{w \text{ in } (x-\xi, x)} \left|\frac{\cos \varepsilon w}{\varepsilon w} - \frac{\sin \varepsilon w}{\varepsilon^2 w^2}\right| \qquad (22.321) \\ &\leqq 2\varepsilon|\xi| \max \frac{1}{|\varepsilon w|} \\ &\leqq \left|\frac{8\xi}{|x|+|\xi|}\right|.\end{aligned}$$

Thus and hence ly an application of (22.32)

$$\frac{1}{\varepsilon\xi}\int_{-A}^{A} f(x-\xi)\left[\frac{\sin \varepsilon x}{x} - \frac{\sin \varepsilon(x-\xi)}{x-\xi}\right] e^{iux}\, dx \tag{22.33}$$

converges in the mean uniformly in u as $A \to \infty$ over any finite range of ξ. Hence by our lemma, we may write

$$\begin{aligned}&s_1(u+\varepsilon) - s_1(u-\varepsilon) - (s(u+\varepsilon) - s(u-\varepsilon)) \int_{-\infty}^{\infty} K_1(\xi)\, e^{iu\xi}\, d\xi \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty} K_1(\xi)\, d\xi \operatorname*{l.i.m.}_{A\to\infty} \int_{-A}^{A} f(x-\xi)\left[\frac{\sin \varepsilon x}{x} - \frac{\sin \varepsilon(x-\xi)}{x-\xi}\right] e^{iux}\, dx.\end{aligned} \tag{22.34}$$

By a further use of (22.321) and our lemma, since

$$\operatorname*{l.i.m.}_{\varepsilon\to 0} \frac{1}{\varepsilon^{1/2}\xi} \operatorname*{l.i.m.}_{A\to\infty} \frac{1}{\pi}\int_{-A}^{A} f(x-\xi)\left[\frac{\sin \varepsilon x}{x} - \frac{\sin \varepsilon(x-\xi)}{x-\xi}\right] e^{iux}\, dx = 0 \tag{22.35}$$

uniformly in ξ, it follows that

$$\begin{aligned}\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\int_{-\infty}^{\infty} \Big| &s_1(u+\varepsilon) - s_1(u-\varepsilon) \\ &- (s(u+\varepsilon) - s(u-\varepsilon)) \int_{-\infty}^{\infty} K_1(\xi)\, e^{iu\xi}\, d\xi\Big|^2 du = 0.\end{aligned} \tag{22.36}$$

From (22.12) and a theorem of the previous paper,[59] it follows that

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\int_{-\infty}^{\infty} |s_1(u+\varepsilon) - s_1(u-\varepsilon)|^2\, du = 0. \tag{22.37}$$

[59] Wiener (7), (5.52).

Combining (22.36) and (22.37), we get

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{-\infty}^{\infty}\left|\int_{-\infty}^{\infty}K_1(\xi)\,e^{iu\xi}\,d\xi\right|^2|s(u+\varepsilon)-s(u-\varepsilon)|^2\,du=0. \tag{22.38}$$

Since over any finite range, by (22.16),

$$\left|\int_{-\infty}^{\infty}K_1(\xi)\,e^{iu\xi}\,d\xi\right|>\text{const.}, \tag{22.39}$$

it results that for any finite C,

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{-C}^{C}|s(u+\varepsilon)-s(u-\varepsilon)|^2\,du=0. \tag{22.40}$$

As in (22.36), we have

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{-\infty}^{\infty}\left|s_2(u+\varepsilon)-s_2(u-\varepsilon)-(s(u+\varepsilon)-s(u-\varepsilon))\int_{-\infty}^{\infty}K_2(\xi)\,e^{iu\xi}\,d\xi\right|^2du=0. \tag{22.41}$$

This yields us

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{-C}^{C}\left|s_2(u+\varepsilon)-s_2(u-\varepsilon)-(s(u+\varepsilon)-s(u-\varepsilon))\int_{-\infty}^{\infty}K_2(\xi)\,e^{iu\xi}\,d\xi\right|^2du=0. \tag{22.42}$$

Combining this with (22.40), we get

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{-C}^{C}\left|s_2(u+\varepsilon)-s_2(u-\varepsilon)\right|^2du=0. \tag{22.43}$$

By our previous paper,[60]

$$\frac{1}{\pi\varepsilon}\int_{-\infty}^{\infty}|f_2(x)|^2\frac{\sin^2\varepsilon x}{x^2}\,dx=\frac{1}{2\varepsilon}\int_{-\infty}^{\infty}|s_2(u+\varepsilon)-s_2(u-\varepsilon)|^2\,du, \tag{22.44}$$

and

$$\frac{1}{\pi\varepsilon}\int_{-\infty}^{\infty}\frac{\sin^2\varepsilon x}{x^2}\,dx\left|\frac{1}{2\lambda}\int_{-\lambda}^{\lambda}f_2(x+y)\,dy\right|^2=\frac{1}{2\varepsilon}\int_{-\infty}^{\infty}\frac{\sin^2\lambda u}{\lambda^2u^2}|s_2(u+\varepsilon)-s_2(u-\varepsilon)|^2\,du. \tag{22.45}$$

[60] Wiener (7), (5.52).

Now

$$\overline{\lim_{\varepsilon\to 0}}\left|\frac{1}{\pi\varepsilon}\int_{-\infty}^{\infty}|f_2(x)|^2\frac{\sin^2\varepsilon x}{x^2}\,dx-\frac{1}{\pi\varepsilon}\int_{-\infty}^{\infty}\frac{\sin^2\varepsilon x}{x^2}\,dx\left|\frac{1}{2\lambda}\int_{-\lambda}^{\lambda}f_2(x+y)\,dy\right|^2\right|$$

$$\leqq \overline{\lim_{\varepsilon\to 0}}\frac{\text{const.}}{\varepsilon}\int_{-\infty}^{\infty}\left|f_2(x)-\frac{1}{2\lambda}\int_{-\lambda}^{\lambda}f_2(x+y)\,dy\right|^2\frac{\sin^2\varepsilon x}{x^2}\,dx$$

$$= \overline{\lim_{\varepsilon\to 0}}\frac{\text{const.}}{\varepsilon}\int_{-\infty}^{\infty}\left|\int_{-\infty}^{\infty}f(x-\xi)\right.$$

$$\left.\times\left[K_2(\xi)-\frac{1}{2\lambda}\int_{-\lambda}^{\lambda}K_2(\xi+y)\,dy\right]d\xi\right|^2\frac{\sin^2\varepsilon x}{x^2}\,dx$$

$$\text{(22.46)}\quad \leqq \overline{\lim_{B\to\infty}}\,\text{const.}\frac{1}{2B}\int_{-B}^{B}\left|\int_{-\infty}^{\infty}f(x-\xi)\right.$$

$$\left.\times\left[K_2(\xi)-\frac{1}{2\lambda}\int_{-\lambda}^{\lambda}K_2(\xi+y)\,dy\right]d\xi\right|^2 dx$$

$$\leqq \text{const.}\,\overline{\lim_{B\to\infty}}\,\frac{1}{B}\int_{-\infty}^{\infty}\left|K_2(\xi)-\frac{1}{2\lambda}\int_{-\lambda}^{\lambda}K_2(\xi+y)\,dy\right|d\xi$$

$$\times\int_{-\infty}^{\infty}\left|K_2(\eta)-\frac{1}{2\lambda}\int_{-\lambda}^{\lambda}K_2(\eta+z)\,dz\right|d\eta\,(B+|\xi|+|\eta|)$$

$$\leqq \text{const.}\,\overline{\lim_{B\to\infty}}\left[\int_{-\infty}^{\infty}\left|K_2(\xi)-\frac{1}{2\lambda}\int_{-\lambda}^{\lambda}K_2(\xi+y)\,dy\right|\left(1+\frac{|\xi|}{B}\right)d\xi\right]^2,$$

where

$$\text{(22.47)}\quad \overline{\lim_{B\to\infty}}\int_{-\infty}^{\infty}\left|K_2(\xi)-\frac{1}{2\lambda}\int_{-\lambda}^{\lambda}K_2(\xi+y)\,dy\right|\left(1+\frac{|\xi|}{B}\right)d\xi$$

$$=\int_{-\infty}^{\infty}\left|K_2(\xi)-\frac{1}{2\lambda}\int_{-\lambda}^{\lambda}K_2(\xi+y)\,dy\right|d\xi.$$

Hence

$$\overline{\lim_{\varepsilon\to 0}}\left|\frac{1}{\pi\varepsilon}\int_{-\infty}^{\infty}|f_2(x)|^2\frac{\sin^2\varepsilon x}{x^2}\,dx-\frac{1}{\pi\varepsilon}\int_{-\infty}^{\infty}\frac{\sin^2\varepsilon x}{x^2}\,dx\left|\frac{1}{2\lambda}\int_{-\lambda}^{\lambda}f_2(x+y)\,dy\right|^2\right|$$

$$\text{(22.48)}\quad \leqq \text{const.}\left[\int_{-\infty}^{\infty}\left|K_2(\xi)-\frac{1}{2\lambda}\int_{-\lambda}^{\lambda}K_2(\xi+y)\,dy\right|d\xi\right]^2.$$

As we have seen in section 1,

$$\text{(22.49)}\quad \lim_{\lambda\to 0}\int_{-\infty}^{\infty}\left|K_2(\xi)-\frac{1}{2\lambda}\int_{-\lambda}^{\lambda}K_2(\xi+y)\,dy\right|d\xi=0,$$

from which it immediately follows that

$$\lim_{\lambda\to 0}\overline{\lim_{\varepsilon\to 0}}\left|\frac{1}{\pi\varepsilon}\int_{-\infty}^{\infty}|f_2(x)|^2\frac{\sin^2\varepsilon x}{x^2}\,dx\right. \tag{22.50}$$
$$\left.-\frac{1}{\pi\varepsilon}\int_{-\infty}^{\infty}\frac{\sin^2\varepsilon x}{x^2}\,dx\left|\frac{1}{2\lambda}\int_{-\lambda}^{\lambda}f_2(x+y)\,dy\right|^2\right| = 0.$$

In combination with (22.44) and (22.45) this yields us

$$\lim_{\lambda\to 0}\overline{\lim_{\varepsilon\to 0}}\frac{1}{2\varepsilon}\int_{-\infty}^{\infty}\left(1-\frac{\sin^2\lambda u}{\lambda^2u^2}\right)|s_2(u+\varepsilon)-s_2(u-\varepsilon)|^2\,du = 0. \tag{22.51}$$

Since

$$1-\frac{\sin^2\lambda u}{\lambda^2u^2}>\frac{1}{2}\quad\text{if}\quad u>\frac{2}{\lambda} \tag{22.52}$$

it results at once that

$$\lim_{C\to\infty}\overline{\lim_{\varepsilon\to 0}}\frac{1}{2\varepsilon}\left[\int_{C}^{\infty}+\int_{-\infty}^{-C}\right]|s_2(u+\varepsilon)-s_2(u-\varepsilon)|^2\,du = 0. \tag{22.53}$$

Combining this with (22.43), we may readily conclude that

$$\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_{-\infty}^{\infty}|s_2(u+\varepsilon)-s_2(u-\varepsilon)|^2\,du = 0, \tag{22.54}$$

or in view of a theorem of our previous paper, that

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}|f_2(x)|^2\,dx = 0. \tag{22.13}$$

This establishes Theorem XXIV. It will be noted that this theorem has an immediate extension in the direction of Theorem XI′.

There is another Tauberian theorem concerning strong summability which we may discuss here. It is due to Hardy and Littlewood.[61] The particular case of it which most directly interests us is the following:

THEOREM XXV. *Let*

$$\frac{1}{\xi}\int_{0}^{\xi}|f(x)|^2\,dx \tag{22.55}$$

be bounded, and let

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{0}^{\varepsilon}f(x)\,dx = 0. \tag{22.56}$$

[61] Hardy and Littlewood (16).

Then if

$$s_n(0) = \frac{1}{\pi}\int_0^\pi f(x)\,\frac{\sin\frac{2n+1}{2}x}{\sin\frac{x}{2}}\,dx \tag{22.57}$$

it follows that

$$\lim_{N\to\infty}\frac{1}{N}\sum_0^N |s_n(0)|^2 = 0. \tag{22.58}$$

In the proof of this theorem, it is easy to see that we may replace the conclusion (22.58) which we are to establish, by the equivalent

$$\lim_{H\to\infty}\frac{1}{H}\int_0^H\left|\frac{2}{\pi}\int_0^\pi f(x)\,\frac{\sin ux}{x}\,dx\right|^2 du = 0. \tag{22.59}$$

By a Tauberian theorem of the type of Theorem XI, already established in a previous paper of the author,[58] it follows that (22.59) is completely equivalent to

$$\lim_{\varepsilon\to 0}\frac{2}{\pi}\int_0^\infty \frac{\sin^2\varepsilon u}{\varepsilon u^2}\left|\frac{2}{\pi}\int_0^\pi f(x)\frac{\sin ux}{x}\,dx\right|^2 du. \tag{22.60}$$

On the assumption that $f(x) = 0$ if $x > \pi$, we have by the Plancherel theorem

$$\begin{aligned}
&\frac{2}{\pi}\int_0^\infty \frac{\sin^2\varepsilon u}{\varepsilon u^2}\left|\frac{2}{\pi}\int_0^\pi f(x)\frac{\sin ux}{x}\,dx\right|^2 du\\
&= \frac{2}{\pi\varepsilon}\int_0^\infty\left|\frac{2}{\pi}\int_0^\pi f(x)\,\frac{\sin\varepsilon u\sin ux}{ux}\,dx\right|^2 du\\
&= \frac{2}{\pi\varepsilon}\int_0^\infty\left|\frac{2}{\pi}\int_0^\pi \frac{\sin\varepsilon u\sin ux}{u}\,d\int_x^\infty\frac{f(\xi)}{\xi}\,d\xi\right|^2 du\\
&= \frac{2}{\pi\varepsilon}\int_0^\infty\left|\frac{2}{\pi}\int_0^\infty \sin\varepsilon u\cos ux\,dx\int_x^\infty\frac{f(\xi)}{\xi}\,d\xi\right|^2 du\\
&= \frac{2}{\pi\varepsilon}\int_0^\infty\left|\frac{1}{\pi}\int_0^\infty(\sin u(x+\varepsilon)-\sin u(x-\varepsilon))dx\int_x^\infty\frac{f(\xi)}{\xi}d\xi\right|^2 du \qquad (22.61)\\
&= \frac{2}{\pi\varepsilon}\int_0^\infty\left|\frac{1}{2\pi}\int_{-\infty}^\infty(\sin u(x+\varepsilon)-\sin u(x-\varepsilon))dx\int_{|x|}^\infty\frac{f(\xi)}{\xi}d\xi\right|^2 du\\
&= \frac{2}{\pi\varepsilon}\int_0^\infty\left|\frac{1}{2\pi}\int_{-\infty}^\infty \sin ux\,dx\int_{|x-\varepsilon|}^{|x+\varepsilon|}\frac{f(\xi)}{\xi}\,d\xi\right|^2 du\\
&= \frac{1}{\pi^2\varepsilon}\int_0^\infty\left|\int_{|x-\varepsilon|}^{|x+\varepsilon|}\frac{f(\xi)}{\xi}\,d\xi\right|^2 dx.
\end{aligned}$$

Thus (22.59) and (22.58) are equivalent to

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_0^\infty\left|\int_{|x-\varepsilon|}^{|x+\varepsilon|}\frac{f(\xi)}{\xi}\,d\xi\right|^2 dx = 0. \tag{22.62}$$

By the Schwarz inequality,

$$\begin{aligned}\frac{1}{\varepsilon}\left|\int_{|x-\varepsilon|}^{|x+\varepsilon|}\frac{f(\xi)}{\xi}\,d\xi\right|^2 &\leqq \frac{1}{\varepsilon}\int_{|x-\varepsilon|}^{|x+\varepsilon|}|f(\xi)|^2\,d\xi\int_{|x-\varepsilon|}^{|x+\varepsilon|}\frac{d\xi}{\xi^2}\\ &= \left(\frac{1}{|x-\varepsilon|}-\frac{1}{|x+\varepsilon|}\right)\frac{1}{\varepsilon}\int_{|x-\varepsilon|}^{|x+\varepsilon|}|f(\xi)|^2\,d\xi.\end{aligned} \tag{22.63}$$

Thus by the boundedness of (22.55) we have uniformly in ε for all sufficiently large values of N:

$$\begin{aligned}\frac{1}{\varepsilon}\int_{N\varepsilon}^\infty\left|\int_{|x-\varepsilon|}^{|x+\varepsilon|}\frac{f(\xi)}{\xi}\,d\xi\right|^2 dx &\leqq \int_{N\varepsilon}^\infty\frac{2\varepsilon}{x^2-\varepsilon^2}\,d\int_0^x\frac{d\eta}{\varepsilon}\int_{|\eta-\varepsilon|}^{|\eta+\varepsilon|}|f(\xi)|^2\,d\xi\\ &\leqq \int_{N\varepsilon}^\infty\frac{4\varepsilon x\,dx}{(x^2-\varepsilon^2)^2}\int_0^x\frac{d\eta}{\varepsilon}\int_{|\eta-\varepsilon|}^{|\eta+\varepsilon|}|f(\xi)|^2\,d\xi\\ &= \frac{1}{2}\int_{N\varepsilon}^\infty\frac{4\varepsilon x\,dx}{(x^2-\varepsilon^2)^2}\int_{-x-\varepsilon}^{x+\varepsilon}|f(\xi)|^2\,d\xi\int_{\xi-\varepsilon}^{\xi+\varepsilon}\frac{d\eta}{\varepsilon}\\ &\leqq \text{const.}\int_{N\varepsilon}^\infty\frac{4\varepsilon x^2\,dx}{(x^2-\varepsilon^2)^2}\\ &= \text{const.}\int_N^\infty\frac{4x^2\,dx}{(x^2-1)^2}.\end{aligned} \tag{22.64}$$

This can be made as small as we wish by taking N large enough. Moreover,

$$\begin{aligned}\frac{1}{\varepsilon}\int_0^{N\varepsilon}\left|\int_{|x-\varepsilon|}^{|x+\varepsilon|}\frac{f(\xi)}{\xi}\,d\xi\right|^2 &= \frac{1}{\varepsilon}\int_0^{N\varepsilon}\left|\int_{|x-\varepsilon|}^{|x+\varepsilon|}\frac{1}{\xi}\,d\int_0^\xi f(\eta)\,d\eta\right|^2 dx\\ &= \frac{1}{\varepsilon}\int_0^{N\varepsilon}\left|\left[\frac{1}{\xi}\int_0^\xi f(\eta)\,d\eta\right]_{|x-\varepsilon|}^{|x+\varepsilon|}+\int_{|x-\varepsilon|}^{|x+\varepsilon|}\frac{d\xi}{\xi^2}\int_0^\xi f(\eta)\,d\eta\right|^2 dx\\ &\leqq \frac{1}{\varepsilon}\int_0^{N\varepsilon}\left|A\left(1+\log\frac{|x+\varepsilon|}{|x-\varepsilon|}\right)\right|^2 dx\\ &= A^2\int_0^N\left(1+\log\frac{|x+1|}{|x-1|}\right)^2 dx,\end{aligned} \tag{22.65}$$

where

$$A > \left|\frac{1}{\xi}\int_0^\xi f(\eta)\,d\eta\right| \qquad (0<\xi\leqq|N+1|\varepsilon) \tag{22.651}$$

By (22.56), we may make A as small as we wish for a given N by taking ε small enough. Now let N be so large that for all ε,

$$\frac{1}{\varepsilon}\int_{N\varepsilon}^{\infty}\left|\int_{|x-\varepsilon|}^{|x+\varepsilon|}\frac{f(\xi)}{\xi}\,d\xi\right|^2 dx < \varepsilon_1/2, \tag{22.66}$$

and then let ε be so small that

$$\frac{1}{\varepsilon}\int_{0}^{N\varepsilon}\left|\int_{|x-\varepsilon|}^{|x+\varepsilon|}\frac{f(\xi)}{\xi}\,d\xi\right|^2 dx < \varepsilon_1/2. \tag{22.67}$$

It will then follow that for this and all smaller values of

$$\frac{1}{\varepsilon}\int_{0}^{\infty}\left|\int_{|x-\varepsilon|}^{|x+\varepsilon|}\frac{f(\xi)}{\xi}\,d\xi\right|^2 dx < \varepsilon_1. \tag{22.68}$$

Thus (22.62) follows, and Theorem XXV is established.

23. **Generalized types of summability.** The subject-matter of the last section leads us to interesting reflections on the notion of summability itself. The ordinary processes for summing a series are linear processes: that is, they consist in replacing the partial sums of a series by linear combinations of partial sums, and then investigating the ordinary limits of these linear combinations. This much of linearity must always remain in a definition of summability, that if two summable series are added term by term, the corresponding sums are also added, in the sense that the new sum-series will be summable to the sum of the sums corresponding to the individual series.

It is possible, however, to consider summability from another standpoint, from which linearity is not so obvious an attribute of the process. We may confine our attention to series, or rather to their sequences of partial sums, which are summable to zero. Thus a method of summability will sum to A the series whose partial sums are:

$$s_1, s_2, \cdots, s_n, \cdots \tag{23.01}$$

if it will sum to zero the series whose partial sums are:

$$s_1 - A, s_2 - A, \cdots, s_n - A, \cdots. \tag{23.02}$$

We may thus center our concept of summability about the notion of *null-sequence* or *null-function*. Whatever definition we choose for the class of such sequences or functions, this class should be closed additively, in the sense that the sum of any two members of the class must belong to

the class. *It is not however essential that a member of the class should be characterized by the vanishing of some particular linear transform of that member.* In this sense, we are introducing a non-linear theory of summability.[62]

It is clearly a desideratum of a definition of a null-function that every function tending to a zero limit at infinity should be null. This requirement at once makes the corresponding definition of summability consistent with convergence and inclusive of it in scope. It is not satisfied by definitions of the type that assert that f is null if

$$\int_{-\infty}^{\infty} |f(x)|^p \, dx < \infty \tag{23.03}$$

but is satisfied by the definition that f is null if

$$\lim_{B\to\infty} \frac{1}{B} \int_{-B}^{B} |f(x)|^2 dx = 0. \tag{23.04}$$

We have already seen that if $s(u)$ is defined as in (22.23), this last statement is equivalent to the assertion that

$$\lim_{\varepsilon\to 0} \frac{1}{2\varepsilon} \int_{-\infty}^{\infty} |s(u+\varepsilon) - s(u-\varepsilon)|^2 du = 0. \tag{23.05}$$

This suggests an even looser definition of a null-function, according to which a function $f(x)$ is null if

$$\frac{1}{2B} \int_{-B}^{B} |f(x)|^2 dx \tag{23.06}$$

is bounded, and if

$$\lim_{\varepsilon\to 0} \frac{1}{2\varepsilon} \int_{-A}^{A} |s(u+\varepsilon) - s(u-\varepsilon)|^2 du = 0 \tag{23.07}$$

for all finite A, or even if (23.07) is true for *some* A greater than 0. The first of these definitions will make $f(x)$ null if (23.06) holds and

$$\int_{-\infty}^{\infty} K(x-\xi) f(\xi) \, d\xi \tag{23.08}$$

is null, provided that

$$\int_{-\infty}^{\infty} (1+|x|)^{3+\varepsilon} |K_1(x)|^2 \, dx < \infty \tag{23.09}$$

and

$$\int_{-\infty}^{\infty} K(x) \, e^{iux} \, dx \neq 0, \qquad [-\infty < u < \infty]. \tag{23.10}$$

[62] As far as the author knows, strong summability of various sorts is the only example of a summability process of this sort now in the literature.

The second definition will have the same property, provided only (23.10) holds for all values of u in some neighborhood of 0. All this follows from the theory developed in the last section.

These two last definitions of summability therefore fit in well with the ordinary linear definitions of summability, which define the generalized limit of a function as an expression of the form

$$\lim_{x\to\infty} \frac{\int_{-\infty}^{\infty} K(x,\xi) f(\xi)\, d\xi}{\int_{-\infty}^{\infty} K(x,\xi)\, d\xi} \tag{23.11}$$

on an appropriate scale of measurement. They have the great advantage of not involving any reference to a particular kernel $K(x, \xi)$. They are, however, restricted to functions for which (23.06) is bounded.

If we confine our attention to functions $f(x)$ vanishing for negative arguments, this difficulty may be overcome in its turn. We now put

$$\sigma(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)\, e^{zx}\, dx \tag{23.12}$$

throughout the half-plane $\Re(z) > u_0$: Then, by analytic continuation if necessary, $\sigma(iv)$ may be defined in some cases on a section of the real axis containing the origin. We put

$$s(u) = \int_0^u \sigma(iw)\, dw \tag{23.121}$$

giving σ its boundary value along the axis of imaginaries. We now define nullity as in (23.06) and (23.07). If for all z with real part between $-\varepsilon$ and $u_0+\varepsilon$

$$\int_{-\infty}^{\infty} |K(x)|\, e^{zx}\, dx < \infty, \tag{23.13}$$

we have

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{zx}\, dx \int_{-\infty}^{\infty} K(x-\xi) f(\xi)\, d\xi = \sigma(z)\int_{-\infty}^{\infty} K(\xi)\, e^{z\xi}\, d\xi, \tag{23.14}$$

so that if $\int_{-\infty}^{\infty} K(\xi)\, e^{z\xi}\, d\xi$ is free from zeros and singularities in the left half plane, $f(x)$ and $\int_{-\infty}^{\infty} K(x-\xi) f(\xi)\, d\xi$ are null or not null simultaneously. With an appropriate definition of "analytic continuation", zeros and singularities not at the origin are of no concern, provided the function

$\int_{-\infty}^{\infty} K(\xi)\, e^{s\xi}\, d\xi$ is representible on a single-sheeted Riemann surface in the left half-plane.

It will be noted that all the definitions here given of the nullity of $f(x)$ depend on the scale chosen for x. If for all large x,

$$\frac{y'(x)}{y(A)} < \text{const. for } x < A \tag{23.15}$$

and

$$\frac{1}{2A}\int_{-A}^{A} |f(y(x))|^2\, dx \to 0 \tag{23.16}$$

then it at once follows that

$$\frac{1}{2B}\int_{-B}^{B} |f(x)|^2\, dx \to 0. \tag{23.17}$$

A similar study of the effect of a change of scale on nullity of other types would be of interest.

24. **Some unsolved problems.** In a piece of work of the ambitious length of the present, it is perhaps worth while to point out to the reader promising future directions of research. The following remarks may therefore not be amiss.

(1) The closure of the set of translations of a given function has been investigated in class L_1 and in class L_2. The methods of proof have been widely different in the two cases, but both results may be stated in a single formulation, that the set of translations of $f(x)$ is closed in the appropriate class when and only when the Fourier transform of $f(x)$, when properly defined and chosen, has no zeros. This formulation continues to constitute a reasonable proposition for L_p, where p is intermediate between 1 and 2, or even exterior to this interval. Is this proposition true? Certainly, for neither of the special cases already given is the method of proof extensible without serious modification. My own suspicion is that the general theorem is at least true for $1 \leqq p \leqq 2$.

(2) Obviously the power of Tauberian theorems in number theory has not been exhausted. Is there any Tauberian theorem which will reach from one complex ordinate to another, and enable us to handle the more refined forms of the prime number theorem?

(3) In particular, can we make a direct study of the closure of the set of all polynomials in the functions

$$e^{-\lambda(x-\xi)}\frac{d}{dx}\frac{e^{x-\xi}}{e^{e^{x-\xi}}-1}, \tag{24.01}$$

and thus attack directly the problem of the zeros of the zeta function in the critical strip?

(4) In section 22, the conditions to which $K_1(x)$ and $K_2(x)$ are subjected in Theorem XXIV are probably needlessly stringent. What is the best possible theorem in this connection?

(5) The "quasi-Tauberian" theorem of section 21 has not yet been exploited to the full. In particular, in the discussion of the relations between Riesz summability, Cesàro summability, Hölder summability, and the like, it is extremely desirable to have theorems of this type in which the kernels are not of the form $K(x-y)$, but are in some sense *nearly* of this form. Here the cruder theorems, depending solely on the use of dominant functions, are probably not difficult to elicit. On the other hand, the more refined ones will almost certainly require the full armament of Carleman's theory of singular integral equations and of the modern von Neumann-Stone[63] calculus of operators. Indeed, the time will soon come when the entire Tauberian theory must be reconsidered from the point of view of this calculus.

BIBLIOGRAPHY.

The present list of memoirs, while it cannot in the nature of things be complete, represents an attempt to bring together in one place as much as possible of the literature of Tauberian theorems. Up to 1915, we have been able to draw on the bibliography of the Cambridge tract by Hardy and M. Riesz on The General Theory of Dirichlet Series, and up to about 1924, on Smail's History and Synopsis of the Theory of Summable Infinite Processes. The bibliography also contains certain memoirs mentioned in the notes which are not of themselves Tauberian.

K. Ananda-Rau.

(1) A note on a theorem of Mr. Hardy's. Proc. Lond. Math. Soc., (2) 17 (1918), 334–336.

(2) On Lambert's series. Proc. Lond. Math. Soc., (2) 19 (1921), 1–20.

(3) On the relation between the convergence of a series and its summability by Cesàro's means. Jour. Indian Math. Soc., 15 (1923–5), 264–268.

(4) On Dirichlet's series with positive coefficients. Rendiconti di Palermo, 54 (1930), 455–462.

S. Bochner and G. H. Hardy.

(1) Note on two theorems of Norbert Wiener. Jour. Lond. Math. Soc., 1 (1926), 240–242.

E. Bortolotti.

(1) Sulle condizioni di applicabilità e di coerenza dei processi di sommazione asintotica di algoritmi infiniti. Rend. del R. Ac. di Bologna, 31 (1926–7), 33–41.

[63] Von Neumann (1); Stone (1).

S. Bosanquet.

(1) The summability of Fourier series. Math. Gaz., 15 (1931), 293–296.
(2) On the summability of Fourier series. Proc. Lond. Math. Soc., (2) 31 (1930), 144–164.

L. S. Bosanquet and E. H. Linfoot.

(1) On the zero order summability of Fourier series. Jour. Lond. Math. Soc., 6 (1931), 117–126.
(2) Generalized means and the summability of Fourier series. Quarterly Journal (Oxford series), 2 (1931), 207–229.

H. E. Bray.

(1) Elementary properties of the Stieltjes integral. Ann. of Math., 20 (1919), 176–186.

T. J. I'a. Bromwich.

(1) An introduction to the theory of infinite series. London 1908. Cf. especially p. 251
(2) On the limits of certain infinite series and integrals. Math. Ann. 65, (1908), 350–369.

Anna Caldarera.

(1) Su talune estenzioni dei criteri di convergenza di Hardy e Landau. Note e memorie di mat. (Catania), 2 (1923), 77–98.

T. Carleman.

(1) A theorem concerning Fourier series. Proc. Lond. Math. Soc., (2) 21 (1923), 483–492.

H. S. Carslaw.

(1) Introduction to the theory of Fourier series and integrals. New edition London 1921. Cf. especially pp. 150–154, 234–243.

S. Chapman.

(1) On non-integral orders of summability of series and integrals. Proc. Lond. Math. Soc., (2) 9 (1910–11), 369–409. Cf. p. 374, second note.

M. Cipolla.

(1) Criteri di convergenza riducibili a quello di Hardy-Landau. Naples Rendiconti (3a), 27 (1921), 28–37.

P. J. Daniell.

(1) A general form of integral. Ann. of Math., 19 (1917), 279–294.

P. Dienes.

(1) Sur la sommabilité de la série de Taylor. C. R., 153 (1911), 802–805.

G. Doetsch.

(1) Ein Konvergenzkriterium für Integrale. Math. Ann., 82 (1920), 68–82.
(2) Die Integrodifferentialgleichungen vom Faltungstypus. Math. Ann., 89 (1923), 192–207.
(3) Sätze von Tauberschem Charakter im Gebiet der Laplace- und Stieltjes Transformation. Berliner Berichte, Phys.-Math. Klasse (1930), 144–157.
(4) Über den Zusammenhang zwischen Abelscher und Borelscher Summabilität. Math. Ann., 104 (1931), 403–414.

P. Fatou.

(1) Séries trigonométriques et séries de Taylor. Acta Math., 30 (1906), 335–400. (Thèse, Paris 1907.)

L. Fejér.

(1) Fourierreihe und Potenzreihe. Monatshefte f. Math., 28 (1917), 64–76.

M. Fekete.

(1) Viszgálatok a Fourier sorokrol. Math. és Termész. Ért., 34 (1916).

M. Fujiwara.

(1) Über die Verallgemeinerung des Tauberschen Satzes auf Doppelreihen. Science reports of the Tôhoku national university, Sendai, Japan, 8 (1919), 43–50.

(2) Über summierbare Reihen und Integrale. Tôhoku math. jour., 15 (1919), 323–329.

(3) Ein Satz über die Borelsche Summation. Tôhoku math. jour., 17 (1920), 339–343.

J. Hadamard.

(1) Sur la distribution des zéros de la fonction $\zeta(s)$ et ses consequences arithmétiques. Bull. Soc. Math. France, 24 (1896), 199–220.

G. H. Hardy.

(1) Researches in the theory of divergent series and divergent integrals. Quarterly journal, 35 (1904), 22–66.

(2) Theorems relating to the summability and convergence of slowly oscillating series. Proc. Lond. Math. Soc., (2) 8 (1909), 301–320.

(3) On the multiplication of Dirichlet series. Proc. Lond. Math. Soc., (2) 10 (1912), 396–405.

(4) On the summability of Fourier's series. Proc. Lond. Math. Soc., (2) 12 (1913), 365–372.

(5) An extension of a theorem on oscillating series. Proc. Lond. Math. Soc., (2), 12 (1913), 174–180.

(6) Note on Lambert's series. Proc. Lond. Math. Soc., (2) 13 (1913), 192–198.

(7) The application of Abel's method of summation to Dirichlet series. Quarterly journal, 47 (1916), 176–192.

(8) The second theorem of consistency for summable series. Proc. Lond. Math. Soc., (2) 15 (1916), 72–88.

(9) On certain criteria for the convergence of the Fourier series of a continuous function. Messenger of mathematics, 49 (1920), 149–155.

(10) The summability of a Fourier series by logarithmic means. Quarterly Journal (Oxford series), 2 (1931), 107–113.

G. H. Hardy and J. E. Littlewood.

(1) Contributions to the arithmetic theory of series. Proc. Lond. Math. Soc., (2) 11 (1913), 411–478.

(2) The relations between Borel's and Cesàro's methods of summation. Proc. Lond. Math. Soc., (2) 11 (1913), 1–16.

(3) Sur la série de Fourier d'une fonction à carré sommable. C. R. April 28, 1913.

(4) Tauberian theorems concerning power series and Dirichlet's series whose coefficients are positive. Proc. Lond. Math. Soc., (2) 13 (1913), 174–191.

(5) Some theorems concerning Dirichlet's series. Mess. math., 43 (1914), 134–147.

(6) Theorems concerning the summability of series by Borel's exponential method. Rendiconti di Palermo, 41 (1916), 36–53.

(7) Sur la convergence des séries de Fourier et des séries de Taylor. C. R. Dec. 24, 1917.

(8) The Riemann zeta-function and the theory of the distribution of primes. Acta Math., 41 (1918), 119–196.

(9) On the Fourier series of a bounded function. Proc. Lond. Math. Soc., (2) 17 (1918), xiii–xv. (Abstract).

(10) Abel's theorem and its converse. Proc. Lond. Math. Soc., (2) 18 (1920), 205–235.

(11) On a Tauberian theorem for Lambert's series, and some fundamental theorems in the analytic theory of numbers. Proc. Lond. Math. Soc., (2) 19 (1921), 21-29.
(12) Solution of the Cesàro summability problem for power-series and Fourier series. Math. Ztschr., 19 (1923), 67-96.
(13) Abel's theorem and its converse (II). Proc. Lond. Math. Soc., (2) 22 (1924), 254-269.
(14) The allied series of a Fourier series. Proc. Lond. Math. Soc., 24 (1925), 211-246.
(15) The strong summability of Fourier series. Proc. Lond. Math. Soc., (2) 26 (1926), 273-286.
(16) Elementary theorems concerning power series with positive coefficients and moment-constants of positive functions. Crelle, 157 (1926), 141-158.
(17) A further note on the converse of Abel's theorem. Proc. Lond. Math. Soc., (2) 25 (1926), 219-236.
(18) Notes on the theory of series (VII): on Young's convergence criterion for Fourier series. Proc. Lond. Math. Soc., (2) 28 (1928) 301-311.
(19) Notes on the theory of series (XI): on Tauberian theorems. Proc. Lond. Math. Soc., (2) 30 (1929), 23-37.
(20) Notes on the theory of series (III): on the summability of the Fourier series of a nearly continuous function. Proc. Camb. Phil. Soc., 23 (1927), 681-684.
(21) The equivalence of certain integral means. Proc. Lond. Math. Soc. records, (2) 22 (1924), xl-xliii.

G. H. Hardy and M. Riesz.

(1) The general theory of Dirichlet series. Cambridge Tract in Mathematics and Mathematical Physics, No. 18. Cambridge 1915.

G. H. Hardy and E. C. Titchmarsh.

(1) Self-reciprocal functions. Quarterly journal (Oxford series), 1 (1930), 196-232.

W. A. Hurwitz.

(1) A trivial Tauberian theorem. Bull. Am. Math. Soc., 32 (1926), 77-82.

S. Ikehara.

(1) An extension of Landau's theorem in the analytic theory of numbers. Journal of Math. and Phys. of the Mass. Inst. of Technology, 10 (1931), 1-12.

S. Izumi.

(1) A generalization of Tauber's theorem. Proc. Imperial Acad. Japan, 5 (1929), 57-59.

M. Jacob.

(1) Über den Eindeutigkeitssatz in der Theorie der verallgemeinerten trigonometrischen Integrale. Math. Ann., 100 (1926-7), 278-294.
(2) Über ein Theorem von Bochner-Hardy-Wiener. Jour. Lond. Math. Soc., 3 (1928), 182-187.

J. Karamata.

(1) Sur le mode de croissance regulière des fonctions. Mathematica (Cluj, Roumania), 4 (1930), 38-53.
(2) Über die Hardy-Littlewoodschen Umkehrungen des Abelschen Stetigkeitssatzes. Math. Zeitschr., 32 (1930), 319-320.
(3) Neuer Beweis und Verallgemeinerung einiger Tauberian-Sätze. Math. Zeitschr., 33 (1931), 294-300.
(4) Neuer Beweis und Verallgemeinerung der Tauberschen Sätze, welche die Laplacesche und Stieltjessche Transformation betreffen. Crelle, 164 (1931), 27-40.

A. Kienast.

(1) Extension to other series of Abel's and Tauber's theorems on power series. Proc. Lond. Math. Soc., (2) 25 (1926), 45–52.

K. Knopp.

(1) Neuere Untersuchungen in der Theorie der divergenten Reihen. Jahresber. d. deutschen Mathematikervereinigung, 32 (1923), 43–67.

(2) Theorie und Anwendung der unendlichen Reihen. Berlin, Springer, 1922. Chap. XIII.

E. Landau.

(1) Über die Konvergenz einiger Klassen von unendlichen Reihen am Rande des Konvergenzgebietes. Monatsh. f. Math. und Phys., 18 (1907), 8–28.

(2) Handbuch der Verteilung der Primzahlen. Leipzig 1909, 2v.

(3) Über die Bedeutung einiger neuerer Grenzwertsätze der Herren Hardy und Axer. Prace mat.-Fiz., 21 (1910), 97–177.

(4) Über einen Satz des Herrn Littlewood. Rendiconti di Palermo, 35 (1913), 265–276.

(5) Ein neues Konvergenzkriterium für Integrale. Münch. Sitzber., 43 (1913), 461–467.

(6) Darstellung und Begründung einiger neuerer Ergebnisse der Funktionentheorie. Berlin 1916, 2nd ed., 1930.

(7) Sobre los números primos en progressión aritmética. Rev. Mat. Hispano-Americana, 4 (1923), 1–16, 33–44.

(8) Vorlesungen über Zahlentheorie. Leipzig 1927, 3v.

H. Lebesgue.

(1) Recherches sur la convergence des séries de Fourier. Math. Ann., 61 (1905), 251–280.

P. Lévy.

(1) Sur les conditions d'application et sur la régularité des procédés de sommation des séries divergentes. Bull. Soc. Math. de France, 54 (1926), 1–25.

S. B. Littauer.

(1) On a theorem of Jacob. Jour. Lond. Math. Soc., 4 (1929), 226–231.

(2) A new Tauberian theorem with application to the summability of Fourier series and integrals. Jour. Math. Phys. Mass. Insti. Tech., 8 (1928), 216–234.

J. E. Littlewood.

(1) On the converse of Abel's theorem on power series. Proc. Lond. Math. Soc., (2) 9 (1910), 434–448.

H. J. Mellin.

(1) Die Theorie der asymptotischen Reihen vom Standpunkte der reziproken Funktionen und Integrale. Ann. Ac. Sci. Fennicae, (A) 18 (1922), No. 4, 108 p.

G. Mignosi.

(1) Inversione d'un teorema sul rapporto delle medie (Cp) di due serie. Naples Rendiconti (3a), 27 (1921), 17–28.

L. Neder.

(1) Über Taubersche Bedingungen. Proc. Lond. Math. Soc., (2) 23 (1925), 172–184.

J. von Neumann.

(1) Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren. Math. Ann., 102 (1930), 49–131.

R. E. A. C. Paley.

(1) On the Cesàro summability of Fourier series and allied series. Proc. Camb. Phil. Soc., 26 (1930), 173–203.

M. Plancherel.

(1) Contribution à l'étude de la répresentation d'une fonction arbitraire par des intégrales définies. Rendiconti di Palermo, 30 (1910), 289–335.

A. Pringsheim.

(1) Über das Verhalten von Potenzreihen auf dem Konvergenzkreise. Münch. Ber., 30 (1900), 43–100.

(2) Über die Divergenz gewisser Potenzreihen an der Konvergenzgrenze. M. Ber., 31 (1901), 505–524.

(3) Über eine Konvergenzbedingung für unendliche Reihen, die durch iterierte Mittelbildung reduzibel sind. M. Ber., 50 (1920), 275–284.

M. Riesz.

(1) Une methode de sommation équivalente à la methode des moyennes arithmétiques. C. R., 12 June 1911.

(2) Über einen Satz des Herrn Fatou. Crelle, 140 (1911) 89–99.

R. Schmidt.

(1) Über das Borelsche Summierungsverfahren. Schriften der Königsberger gelehrten Gesellschaft, 1 (1925), 202–256.

(2) Über divergente Folgen und lineare Mittelbildungen. M. Ztsch., 22 (1925), 89–152.

W. Schnee.

(1) Über Dirichletsche Reihen. Rend. di Palermo, 27 (1909), 87–116.

J. Schur.

(1) Über lineare Transformationen in der Theorie der unendlichen Reihen. Crelle, 151 (1921), 79–111.

L. L. Smail.

(1) History and synopsis of the theory of summable infinite processes. U. of Oregon publ., 2 (Feb. 1925), No. 8.

M. H. Stone.

(1) Linear transformations in Hilbert space III. Operational methods and group theory. Proc. Mat. Acad. Sci., 16 (1930), 172–175.

O. G. Sutton.

(1) On a theorem of Carleman. Proc. Lond. Math. Soc., (2) 23 (1925), xlvii–li.

O. Szász.

(1) Verallgemeinerung und neuer Beweis einiger Sätze Tauberischer Art. Münch. Ber., (1929), 325–340.

(2) Über einen Satz von Hardy und Littlewood. Berliner Berichte 1930, 4p.

(3) Über die Approximation stetiger Funktionen durch gegebene Funktionenfolgen. Math. Ann., 104 (1930), 155–160.

(4) Abel hatványsortételével kapcsolatos újabb vizsgálatokról. Mat. és Phys. Lapok., 36 (1929), 10–22.

(5) Über Sätze Tauberscher Art. Jahresb. d. Deutschen Mathematikervereinigung, 39 (1930), 28–31.

(6) Über einige Sätze von Hardy und Littlewood. Göttinger Nachrichten (1930), Fachgruppe I, 315–333.

(7) Verallgemeinerung eines Littlewoodschen Satzes über Potenzreihen. J. L. M. S., 3 (1930), 254–262.

(8) Über Dirichletsche Reihen an der Konvergenzgrenze. Rend. del Cong. Math. di Bologna, (1928), 269–276.

A. Tauber.

(1) Ein Satz aus der Theorie der unendlichen Reihen. Monatshefte f. Math., 8 (1897), 273–277.

E. C. Titchmarsh.

(1) A theorem on infinite products. Jour. Lond. Math. Soc., 1 (1926), 35–37.
(2) On integral functions with real negative zeros. Proc. Lond. Math. Soc., (2) 26 (1927), 185–200.
(3) A contribution to the theory of Fourier transforms. Proc. Lond. Math. Soc., (2) 23 (1924), 279–289.

C. de la Vallée Poussin.

(1) Récherches analytiques sur la théorie des nombres premiers. Première partie. Ann. Soc. Sci. Brux., 20: 2 (1896), 183–256.

S. Verblunsky.

(1) The convergence of singular integrals. Proc. Camb. Phil. Soc., 26 (1930), 312–322.
(2) On summable trigonometric series. Proc. Lond. Math. Soc., (2) 31 (1930), 370–386.
(3) On the limit of a function at a point. Proc. Lond. Math. Soc., (2) 32 (1931), 163–200.
(4) A uniqueness theorem for trigonometric series. Quarterly Journal (Oxford series), (2) (1931), 81–90.

T. Vijayaraghavan.

(1) A theorem concerning the summability of a series by Borel's method. Proc. Lond. Math. Soc., (2) 27 (1928), 316–326.
(2) A Tauberian theorem. Jour. Lond. Math. Soc., 1 (1926), 113–230.

N. Wiener.

(1) The harmonic analysis of irregular motion. Jour. Math. and Phys. Mass. Inst. Tech., 5 (1925), 99–122.
(2) Une methode nouvelle pour la démonstration des théorèmes de M. Tauber. C. R., 184 (1927), 793–795.
(3) On a theorem of Bochner and Hardy. Jour. Lond. Math. Soc., 2 (1927), 118–123.
(4) A new method in Tauberian theorems. Jour. Math. Phys. M. I. T., 7 (1928), 161–184.
(5) A type of Tauberian theorem applying to Fourier series. Proc. Lond. Math. Soc., (2) 30 (1929), 1–8.
(6) The spectrum of an arbitrary function. Proc. Lond. Math. Soc., (2) 27 (1928), 487–496.
(7) Generalized harmonic analysis. Acta. Math., 55 (1930), 117–258.
(8) The operational calculus. Math. Ann., 95 (1926), 557–584.

A. Wintner.

(1) Untersuchungen über Funktionen großer Zahlen. (Erste Mitteilung) M. Ztschr., 28 (1928), 416–429.

W. H. Young.

(1) On the determination of the summability of a function by means of its Fourier constants. Proc. Lond. Math. Soc., (2) 12 (1912), 71–88.

SUMMATION OF SERIES AND CONFORMAL MAPPING.[1]

By T. H. Gronwall.

1. **Introduction.** A series $u_0+u_1+u_2+\cdots$ is summable by the method of de la Vallée Poussin, or briefly, summable (V), to the sum s when the expression

$$V_n = \sum_{\nu=0}^{n} \frac{n!\ n!}{(n-\nu)!\ (n+\nu)!}\, u_\nu \tag{1}$$

tends toward s as n increases indefinitely. This method was investigated by its discoverer in the case of Fourier's series only.[2] It was shown by the writer and independently by C. N. Moore[3] that any series summable by Cesàro's means of any order is also summable (V) to the same sum. The original proof involved some rather long and artificial algebraic transformations, and in an effort to avoid these, it was found that with

$$w = \frac{4z}{(1+z)^2} \tag{2}$$

there is the following generating identity[4] for the V_n in (1):

$$\sum_{\nu=0}^{\infty} u_\nu z^\nu = (1-w)^{1/2} \sum_{n=0}^{\infty} \frac{(2n)!}{2^{2n}\, n!\ n!}\, V_n w^n. \tag{3}$$

This is proved in paragraph 2.

By means of (2), the unit circle $|z|<1$ is mapped simply[5] on the w-plane slit along the real axis from 1 to $+\infty$,[6] and consequently $|w|<1$ is mapped

---

[1] Received July 6, 1931. — Part of the results of this paper were presented to the American Mathematical Society, Oct. 31, 1925.

[2] C. J. de la Vallée Poussin, Sur l'approximation des fonctions d'une variable et de leurs dérivées par des polynômes et des suites limitées de Fourier. Bull. Ac. Sc. de Belgique 1908, pp. 193–254.

[3] T. H. Gronwall, Sur quelques méthodes de sommation et leur application à la série de Fourier, Comptes Rendus **158** (1914), pp. 1664–1665. Über einige Summationsmethoden und ihre Anwendung auf die Fouriersche Reihe, Journ. f. Math. **147** (1916), pp. 16–35.

C. N. Moore, Sur la relation entre certaines méthodes pour la sommation d'une série divergente, Comptes Rendus **158** (1914), pp. 1774–1775.

[4] Professor de la Vallée Poussin, to whom the writer communicated this result, was kind enough to include it (without proof) in his Rice Institute lectures: On the approximation of functions of a real variable and on quasi-analytic functions, Rice Institute Pamphlet **12** (1925), pp. 105–172. See p. 117.

[5] i. e. nowhere overlapping (German: schlichte Abbildung).

[6] There is an obvious connection between (2) and the map $w = z/(1-z)^2$, which gives the extremes in the distortion theorem.

simply on a region interior to $|z|<1$, and $z=0$ corresponds to $w=0$. This remark enables us to derive, from simple mapping, a whole class of summation methods, for which the connection with the Cesàro (and Poisson) sums may be established with the same facility as in the case of the summation (V).

The definition of these new summation methods is based on two functions $f(w)$ and $g(w)$ with the following properties:

$$f(w) \text{ is holomorphic for } |w|\leqq 1 \text{ except at } w=1,$$

and

$$z = f(w) \tag{4}$$

maps $|w|<1$ simply on a region D interior to $|z|<1$ in such a manner that $z=0$ corresponds to $w=0$, and $z=1$ to $w=1$. The inverse function is holomorphic on the boundary of D except at $z=1$, and at this point

$$1-w = (1-z)^{\lambda}(a+\cdots), \qquad \lambda \geqq 1, \qquad a>0, \tag{5}$$

where the dots denote a power series in $1-z$ without constant term (the coefficients may be complex). As to $g(w)$, we have

$$g(w) = \sum_{n=0}^{\infty} b_n w^n, \qquad b_n \neq 0, \qquad (n=0,1,2,\cdots), \tag{6}$$

and

$$g(w) = (1-w)^{-\alpha} + \gamma(w), \qquad \alpha>0, \tag{7}$$

where $\gamma(w)$ is holomorphic for $|w|\leqq 1$, and finally

$$g(w) \neq 0 \quad \text{for} \quad |w|<1. \tag{8}$$

DEFINITION. *When $U_0, U_1, U_2, \cdots$ are generated by the identity*

$$\sum_{\nu=0}^{\infty} u_\nu z^\nu = \frac{1}{g(w)} \sum_{n=0}^{\infty} b_n U_n w^n, \tag{9}$$

and

$$U_n \to s \quad \text{as} \quad n\to\infty, \tag{10}$$

then the series $u_0+u_1+u_2+\cdots$ is summable (f, g) to the sum s.

Since $f(0)=0$ in (4), the expansion of z^ν in powers of w begins with the νth power, and it follows from (9) that

$$U_n = \sum_{\nu=0}^{n} a_{n\nu} u_\nu, \qquad (n=0,1,2,\cdots), \tag{11}$$

where the $a_{n\nu}$ depend on the coefficients of $f(w)$ and $g(w)$ only, and not on the u_ν.[7] The Cesàro sums are the limiting case $f(w) = w$, $g(w) = (1-w)^{-k-1}$; here D coincides with $|z|<1$.

The main properties of the (f, g) summation method are contained in the following three theorems, which are proved in paragraphs 3 to 5.

THEOREM 1. (Connection with Poisson summability.) *Let the series* $u_0+u_1+u_2+\cdots$ *be summable* (f, g) *to the sum* s, *so that the function* $\varphi(z)$ *defined in the neighborhood of* $z = 0$ *by*

$$\varphi(z) = \sum_{\nu=0}^{\infty} u_\nu z^\nu$$

is holomorphic inside D. *Then*

$$\varphi(z) \to s \quad \text{as} \quad z \to 1$$

uniformly for every path of z *interior to* D, *which reaches* $z = 1$ *inside the sector*

$$z = 1 - re^{\theta i}, \quad -\theta_0 < \theta < \theta_0, \quad \theta_0 < \pi/2\lambda, \quad r \geqq 0. \tag{12}$$

THEOREM 2. (Connection with Cesàro summability of any order k). *Let* $u_0+u_1+u_2+\cdots$ *be summable* (Ck) *to the sum* s. *When* $\lambda > 1$ *in* (5), *the series is also summable* (f, g) *to the same sum..*[8]

THEOREM 3. (Comparison of two summation methods (f, g)). *Let* $f_1(w)$ *and* $g_1(w)$ *satisfy the same conditions as* $f(w)$ *and* $g(w)$, *but with other constants* λ_1, a_1, b_{1n}, α_1, *and a different map region* D_1. *When* $\lambda_1 > \lambda$ *and* D_1 *is interior to* D, *then any series which is summable* (f, g) *is also summable* (f_1, g_1) *to the same sum.*

In paragraph 6, we introduce a generalized de la Vallée Poussin summation method (Vk) of order $k \geqq 0$, obtained by making

$$\alpha = 2^{-k}, \; f(w) = \frac{1-(1-w)^\alpha}{1+(1-w)^\alpha}, \quad g(w) = (1-w)^{-\alpha}, \tag{13}$$

where $(1-w)^\alpha$ is that branch of the function which reduces to unity at $w = 0$, so that (V) is the special case $k = 1$. The main properties of (Vk), as given by Theorems 1 to 3, are summarized in Theorem 4. We next consider the geometric series $1+\zeta+\zeta^2+\cdots$ and determine the region in which it is summable (Vk) to the sum $1/(1-\zeta)$, Theorem 5; it is shown in particular that for any ζ which is not real and $\geqq 1$, the

[7] When the series to the left in (9) has a zero convergence radius, we may avoid the formal reckoning with divergent series and obtain (11) by making $u_\nu = 0$ for $\nu > n$.

[8] The condition $\lambda > 1$ is essential. As will be shown in the proof of equations (14) in paragraph 7, $\lambda = 1$ for the Euler sums (Ep) of order p introduced by K. Knopp, Über das Eulersche Summierungsverfahren, Math. Zeitschr. **15** (1922), pp. 226–253. Knopp shows by examples, pp. 251–253, that a series may be summable $(C1)$ without being summable (Ep) for any p or, conversely, summable $(E1)$ but not summable (Ck) for any k.

series is summable (Vk) with a sufficiently large k. We finally show that the summations (V, k) and $(V, k+1)$ are connected by a recurrent formula, so that (Vk) takes a particularly simple form when k is restricted to the positive integers.

In paragraph 7, it is shown that the generalized Euler summation (Ep) of order $p \geqq 0$ is obtained with

$$(14)\quad \vartheta = 2^{-p},\ f(w) = \frac{\vartheta w}{1-(1-\vartheta)w},\ g(w) = \frac{\vartheta}{(1-w)\,(1-(1-\vartheta)w)}.$$

Various results due to Knopp[8] are proved, and finally the connection between (Ep) and (Vk) is established, Theorem 6.

2. Proof that the de la Vallée Poussin sums (1) are generated by (3).

We begin by showing that, z being that root of equation (2) which vanishes when $w = 0$, we have for $|w| < 1$

$$(15)\quad (1-w)^{-1/2} z^\nu = \sum_{n=\nu}^{\infty} \frac{(2n)!}{2^{2n}(n-\nu)!\,(n+\nu)!}\, w^n, \qquad (\nu = 0, 1, 2, \cdots).$$

In fact, for $\nu = 0$, (15) is the binomial expansion of $(1-w)^{-1/2}$, and for $\nu = 1$, we observe that (2) gives

$$1-w = \left(\frac{1-z}{1+z}\right)^2,\ z = \frac{1-(1-w)^{1/2}}{1+(1-w)^{1/2}} = \frac{2-w-2(1-w)^{1/2}}{w},$$

whence

$$\begin{aligned}(1-w)^{-1/2} z &= \left(\frac{2}{w}-1\right)(1-w)^{-1/2} - \frac{2}{w} \\ &= \sum_{n=0}^{\infty} \left(\frac{2(2n+2)!}{2^{2n+2}(n+1)!\,(n+1)!} - \frac{(2n)!}{2^{2n} n!\, n!}\right) w^n \\ &= \sum_{n=1}^{\infty} \frac{(2n)!}{2^{2n}(n-1)!\,(n+1)!}\, w^n.\end{aligned}$$

To proceed by complete induction, we note that (2) may be written $z^2 + (2 - 4/w)\, z + 1 = 0$, whence, multiplying by $(1-w)^{-1/2} z^{\nu-1}$,

$$(1-w)^{-1/2} z^{\nu+1} + \left(2 - \frac{4}{w}\right)(1-w)^{-1/2} z^\nu + (1-w)^{-1/2} z^{\nu-1} = 0.$$

Assuming (15) proved up to the exponent $\nu \geqq 1$ inclusive, the preceding equation gives

$$\begin{aligned}(1-w)^{-1/2} z^{\nu+1} &= \left(\frac{4}{w}-2\right) \sum_{n=\nu}^{\infty} \frac{(2n)!}{2^{2n}(n-\nu)!\,(n+\nu)!}\, w^n \\ &\quad - \sum_{n=\nu-1}^{\infty} \frac{(2n)!}{2^{2n}(n-\nu+1)!\,(n+\nu-1)!}\, w^n.\end{aligned}$$

Collecting terms to the right, the coefficient of $w^{\nu-1}$ vanishes, and that of w^n for $n \geqq \nu$ is

$$\frac{4(2n+2)!}{2^{2n+2}(n+1-\nu)!\,(n+1+\nu)!} - \frac{2(2n)!}{2^{2n}(n-\nu)!\,(n+\nu)!} - \frac{(2n)!}{2^{2n}(n-\nu+1)!\,(n+\nu-1)!}$$

$$= \frac{(2n)!\,[(2n+1)(2n+2) - 2(n-\nu+1)(n+\nu+1) - (n+\nu)(n+\nu+1)]}{2^{2n}(n-\nu+1)!\,(n+\nu+1)!}$$

$$= \frac{(2n)!\,(n-\nu)(n-\nu+1)}{2^{2n}(n-\nu+1)!\,(n+\nu+1)!}$$

$$= \begin{cases} 0, & n = \nu; \\ \dfrac{(2n)!}{2^{2n}(n-\nu-1)!\,(n+\nu+1)!}, & n > \nu. \end{cases}$$

Now multiply (3) by $(1-w)^{-1/2}$, introduce (15) and compare coefficients of w^n on both sides; the result is (1).

3. **Proof of Theorem 1.** First we shall establish some preliminaries common to the proofs of Theorems 1 to 3. In the special case $u_0 = s$, $u_\nu = 0$ $(\nu > 0)$, the left member of (9) reduces to s, and consequently $U_n = s$ for all n. Now (11) shows that replacing our series by $(u_0 - s) + u_1 + u_2 + \cdots$, U_n is replaced by $U_n - s$, and consequently there is no loss in generality by assuming $s = 0$. Next, we note that in (6), the exact order of magnitude of b_n is $\alpha - 1$, i. e.

$$b_n = O(n^{\alpha-1}), \qquad n^{\alpha-1} = O(b_n). \tag{16}$$

In fact, the binomial theorem gives

$$(1-w)^{-\alpha} = \sum_{n=0}^{\infty} \frac{\Gamma(n+\alpha)}{\Gamma(\alpha)\,\Gamma(n+1)} w^n = \sum_{n=0}^{\infty} A_n^{(\alpha)} w^n, \tag{17}$$

and using the Cauchy majorant for $\gamma(w)$ in (7), we have

$$|b_n - A_n^{(\alpha)}| < MR^{-n}, \qquad R > 1, \tag{18}$$

whence (16) by the application of Stirling's formula. Finally we show that when w lies in the sector

$$w = 1 - \varrho e^{\varphi i}, \quad -\varphi_0 < \varphi < \varphi_0; \quad \varphi_0 < \pi/2, \quad \varrho > 0, \tag{19}$$

then

$$\sum_{n=0}^{\infty} |b_n|\,|w|^n = O(|g(w)|) \quad \text{as} \quad \varrho \to 0. \tag{20}$$

Since $A_n^{(\alpha)} > 0$ for $\alpha > 0$, we have according to (17) and (18) with $|w| < 1$

$$\sum_{n=0}^{\infty} |b_n|\,|w|^n < \sum_{n=0}^{\infty}\left[A_n^{(\alpha)}|w|^n + M\left(\frac{|w|}{R}\right)^n\right]$$
$$= (1-|w|)^{-\alpha} + MR/(R-|w|),$$

$$|g(w)| > \left|\sum_{n=0}^{\infty} A_n^{(\alpha)} w^n\right| - \sum_{n=0}^{\infty} M\left(\frac{|w|}{R}\right)^n = |1-w|^{-\alpha} - MR/(R-|w|),$$

and (20) follows by the application of the well known inequality[9]

$$\frac{|1-w|}{1-|w|} < \frac{2}{\cos\varphi_0},$$

w in (19) and $\varrho < \cos\varphi_0$.

Turning now to the proof of Theorem 1, we note that by (10) and (16), the series to the right in (9) converges for $|w| < 1$, and by (8) and (4), the power series to the left in (9) converges for $|z|$ sufficiently small and defines a function $\varphi(z)$ holomorphic inside the region D. Consider a path interior to D and reaching $z = 1$ inside the sector (12). We may then choose an angle $\varphi_0 < \pi/2$ so that the corresponding path in the w-plane, obviously interior to $|w| < 1$, reaches $w = 1$ inside the sector (19). As shown above, we may assume $s = 0$, so that to any $\varepsilon_1 > 0$ there corresponds an $N = N(\varepsilon_1)$ such that $|U_n| < \varepsilon_1$ for $n > N$, and consequently for w interior to (19)

$$\left|\sum_{n=0}^{\infty} b_n U_n w^n\right| < \sum_{n=0}^{N} |b_n U_n|\,|w|^n + \sum_{n=N+1}^{\infty} |b_n|\varepsilon_1 |w|^n$$
$$< \sum_{n=0}^{N} |b_n U_n| + \varepsilon_1 \sum_{n=0}^{\infty} |b_n|\,|w|^n.$$

By (7) and (20) there exist two constants c_1 and c_2 such that for w in (19) and ϱ sufficiently small

$$\frac{1}{|g(w)|} < c_1 \varrho^{\alpha}, \qquad \sum_{n=0}^{\infty} |b_n|\,|w|^n < c_2 |g(w)|,$$

and the preceding inequality gives

$$\left|\frac{1}{g(w)} \sum_{n=0}^{\infty} b_n U_n w^n\right| < c_1 \varrho^{\alpha} \sum_{n=0}^{N} |b_n U_n| + c_2 \varepsilon_1. \tag{21}$$

To any assigned $\varepsilon > 0$ we may choose an $\varepsilon_1 < \varepsilon/2c_2$ and then a $\varrho(\varepsilon)$ so small that with $N = N(\varepsilon_1)$, $c_1 \varrho^{\alpha} \sum_{n=0}^{N} |b_n U_n| < \varepsilon/2$ for $\varrho < \varrho(\varepsilon)$. Hence there exists an $r(\varepsilon)$ such that for z in (12) and $r < r(\varepsilon)$

[9] See, for instance, E. Landau, Darstellung und Begründung einiger neuerer Ergebnisse der Funktionentheorie, 2nd ed. (Berlin, Springer, 1929), pp. 54-55.

(22) $$|\varphi(z)|<\varepsilon,$$

and our theorem is proved.

4. **Proof of Theorem 2.** Let the series $u_0+u_1+u_2+\cdots$ be summable (Ck); as before, we may assume the sum to be zero. The generating identity of the Cesàro sums of order k shows that since $s_n^{(k)}\to 0$ as $n\to\infty$,

(23) $$\varphi(z)=\sum_{\nu=0}^{\infty}u_\nu z^\nu=(1-z)^{k+1}\sum_{n=0}^{\infty}A_n^{(k+1)}s_n^{(k)}z^n$$

is holomorphic for $|z|<1$, so that for any $R_1<1$

(24) $$|\varphi(z)|<M_1=M_1(R_1)\quad\text{for}\quad|z|\leqq R_1.$$

Moreover, if in the argument leading to (21) we replace w by z, $g(w)$ by $(1-z)^{-k-1}$, b_n by $A_n^{(k+1)}$ and U_n by $s_n^{(k)}$, we find that to any $\theta_1<\pi/2$ and any $\varepsilon>0$ there corresponds an $r=r(\theta_1,\varepsilon)$ such that (22) holds in the sector

(25) $$z=1-re^{\theta i},\quad -\theta_1<\theta<\theta_1,\quad 0<r<r(\theta_1,\varepsilon).$$

Since we assumed D to be interior to $|z|<1$, and $f(w)$ holomorphic for $|w|\leqq 1$, $w\neq 1$, we may choose any β between 0 and $\pi/2$, and then $\varrho_1>0$ such that $f(w)$ is holomorphic, and $|f(w)|<1$, for

(26) $$|w|\leqq|1+\varrho_1 e^{\beta i}|,\quad |w-1|\geqq\varrho_1.$$

Since $\lambda>1$ by hypothesis, we may choose β and θ_1 such that

$$\frac{\pi-\beta}{\lambda}<\theta_1<\frac{\pi}{2},$$

and in (26), we may make $\varrho_1=\varrho_1(\varepsilon)$ so small that (5) maps the sector

(27) $$1-w=\varrho e^{\varphi i},\quad \beta-\pi\leqq\varphi\leqq\pi-\beta,\quad 0<\varrho\leqq\varrho_1=\varrho_1(\varepsilon)$$

on a part of the z-plane interior to the sector (25). Consequently, we may choose $R_1<1$ such that (24) holds for $z=f(w)$ and w interior to or on the boundary of the region enclosed by the curve $C=C_1+C_2+C_3+C_4$, where

C_1 is the circular arc obtained by making $\varrho=\varrho_0<\varrho_1$ in (27);
C_2 the straight line $w=1+\varrho e^{\beta i}$, $\varrho_0\leqq\varrho\leqq\varrho_1$;
C_3 the circular arc $|w|=|w_1|=|1+\varrho_1 e^{\beta i}|$, $|w-1|\geqq\varrho_1$;
C_4 the straight line $w=1+\varrho e^{-\beta i}$, $\varrho_1\geqq\varrho\geqq\varrho_0$.

Making $z=f(w)$ in (9), the application of Cauchy's integral formula gives

(28) $$2\pi i\,b_n U_n=\int_C\varphi(f(w))\,g(w)\frac{dw}{w^{n+1}}=I_1+I_2+I_3+I_4.$$

On C_1, C_2 and C_4, we have $|\varphi(f(w))|<\varepsilon$ by (22), (25) and (27), and $|g(w)|<c_3\varrho_0^{-\alpha}$ by (7). On C_1, we have $\varrho=\varrho_0$ and $\left|\frac{dw}{w^{n+1}}\right|\leqq\frac{\varrho_0\,d\varphi}{(1-\varrho_0)^{n+1}}$ by (27). Consequently, using (16),

$$\left|\frac{I_1}{2\pi i\,b_n}\right|<\frac{c_4(n\varrho_0)^{1-\alpha}\,\varepsilon}{(1-\varrho_0)^{n+1}}. \tag{29}$$

On C_2 and C_4, we have $|w|^2=1+2\varrho\cos\beta+\varrho^2>1+2\varrho\cos\beta$, so that

$$\begin{aligned}|I_2+I_4|&<2\varepsilon\,c_3\,\varrho_0^{-\alpha}\int_{\varrho_0}^{\varrho_1}(1+2\varrho\cos\beta)^{-\frac{n+1}{2}}\,d\varrho\\&<\frac{\varepsilon\,c_3\,\varrho_0^{-\alpha}}{\cos\beta}\cdot\frac{2}{n-1}(1+2\varrho_0\cos\beta)^{-\frac{n-1}{2}},\end{aligned}$$

and by (16)

$$\left|\frac{I_2+I_4}{2\pi i\,b_n}\right|<c_5(n\varrho_0)^{-\alpha}\,\varepsilon. \tag{30}$$

On C_3, we have $|\varphi(f(w))\,g(w)|<c_6(\varepsilon)$ (dependent on ε but not on n), and

$$|I_3|<c_6(\varepsilon)\left|\int_{w=|w_1|}\frac{dw}{w^{n+1}}\right|\leqq 2\pi\,c_6(\varepsilon)\,|w_1|^{-n},$$

and by (16)

$$\left|\frac{I_3}{2\pi i\,b_n}\right|<c_7(\varepsilon)\,n^{1-\alpha}\,|1+\varrho_1(\varepsilon)\,e^{\beta i}|^{-n}. \tag{31}$$

Now since $0<\beta<\pi/2$ and consequently $|1+\varrho_1(\varepsilon)\,e^{\beta i}|>1$ for $\varepsilon>0$, it follows that by taking first ε sufficiently small, then n sufficiently large and finally $\varrho_0=1/n$, we may make (29), (30), and (31) as small as we please, and hence by (28)

$$U_n\to 0 \quad \text{as} \quad n\to\infty,$$

which proves Theorem 2.

5. **Proof of Theorem 3.** This proof proceeds along the same lines as that of Theorem 2. We have now two mapping functions $z=f(w)$, mapping $|w|<1$ on a simply connected and nowhere overlapping region D in the z-plane, and $z=f_1(w_1)$, mapping $|w_1|<1$ on a region D_1 with the same properties in the z-plane, and D_1 is interior to D. Eliminating z, we obtain a function

$$w=\psi(w_1) \tag{32}$$

mapping $|w_1|<1$ on a simple and nowhere overlapping region Δ *interior to* $|w|<1$; moreover, f_1 being holomorphic on $|w_1|=1$ except at $w_1=1$, and the inverse of f holomorphic on the boundary of D except at $z=1$ it follows that $\psi(w_1)$ is also holomorphic for $|w_1|\leqq 1$ except at $w_1=1$. Eliminating z between (5) and the corresponding equation

$$1-w_1 = a_1(1-z)^{\lambda_1}+\cdots, \qquad \lambda_1 \geqq 1, \qquad a_1>0,$$

we find

$$1-w_1 = a_1\, a^{-\frac{\lambda_1}{\lambda}}(1-w)^{\frac{\lambda_1}{\lambda}}+\cdots. \tag{33}$$

By hypothesis, $U_n \to s$ as $n\to\infty$, and as before, we may assume $s=0$. Consequently

$$\Phi(w) = \frac{1}{g(w)}\sum_{n=0}^{\infty} b_n U_n w^n \tag{34}$$

is holomorphic for $|w|<1$, and as in the proof of Theorem 1, it is seen that for any ε however small,

$$|\Phi(w)|<\varepsilon \tag{35}$$

when w is inside the sector

$$w = 1-re^{\theta i}, \qquad -\theta_1<\theta<\theta_1, \qquad 0<r<r_1(\varepsilon) \tag{25'}$$

where $\theta_1<\pi/2$ is fixed (independent of ε) and $r_1(\varepsilon)$ sufficiently small. Since $\lambda_1>\lambda$ by hypothesis, we may now determine $\beta<\pi/2$ and θ_1 so that

$$\frac{\lambda}{\lambda_1}(\pi-\beta)<\theta_1,$$

and (33) shows that for $\varrho_1(\varepsilon)$ sufficiently small, (32) maps the sector

$$1-w_1 = \varrho e^{\varphi i}, \qquad \beta-\pi \leqq \varphi \leqq \pi-\beta, \qquad 0<\varrho \leqq \varrho_1 = \varrho_1(\varepsilon) \tag{27'}$$

on a part of the w-plane interior to the sector (25'). By (34), (9) and the corresponding equation involving w_1, we have

$$\Phi(w) = \frac{1}{g_1(w_1)}\sum_{n=0}^{\infty} b_{1n} U_{1n} w_1^n,$$

and the proof, by Cauchy's integral formula, that $U_{1n}\to 0$ as $n\to\infty$, is identical with the proof of Theorem 2, except for replacing z by w, $\varphi(z)$ by $\Phi(w)$, w by w_1, $f(w)$ by $\psi(w_1)$ and $g(w)$ by $g_1(w_1)$.

6. **The generalized de la Vallée Poussin summation (Vk) of order $k \geqq 0$.** The nth (Vk) sum we denote by $V_n^{(k)}$, and it is defined as the value of U_n in (9) when $f(w)$ and $g(w)$ are given by (13). The choice of k instead of $\alpha = 2^{-k}$ to denote the order is due to the desire to have the order increase as the summation method grows more powerful (see Theorem 4 below); moreover, the special case when k is restricted to the positive integers is of particularly simple structure on account of the recurrent formula (53).

First, we must verify that (13) satisfies the conditions imposed on $f(w)$ and $g(w)$. With regard to the latter, it is evident that (6) to (8) are fulfilled, and by (17)

$$b_n = b_n^{(\alpha)} = A_n^{(\alpha)} = \frac{\Gamma(n+\alpha)}{\Gamma(\alpha)\,\Gamma(n+1)}. \tag{36}$$

From (4) and (13) we obtain

$$\frac{1-z}{1+z} = (1-w)^{\alpha}, \tag{37}$$

so that (5) is satisfied with

$$\lambda = \frac{1}{\alpha} = 2^k \geqq 1 \tag{38}$$

and $a = 2^{-\lambda}$. To investigate the map of $|w|<1$ it is convenient to introduce the auxiliary mapping function

$$Z = \frac{1-z}{1+z}, \tag{39}$$

so that, by (37)

$$Z = (1-w)^{\alpha}. \tag{40}$$

Now (40) maps the w-plane slit along the real axis from $w = 1$ to $+\infty$, represented in polar coördinates by

$$1 - w = \varrho e^{\varphi i}, \qquad -\pi < \varphi < \pi, \qquad \varrho > 0, \tag{41}$$

upon the sector in the Z-plane

$$Z = r e^{\theta i}, \qquad -\alpha\pi < \theta < \alpha\pi, \qquad r > 0, \tag{42}$$

and the map is simple. In fact, since that branch of $(1-w)^{\alpha}$ was chosen which reduces to 1 at $w = 0$, we have $r = \varrho^{\alpha}$, $\theta = \alpha\varphi$, and two different w-values $\varrho e^{\varphi i}$ and $\varrho' e^{\varphi' i}$ cannot give the same Z unless $\varrho = \varrho'$ and $\alpha\varphi - \alpha\varphi' = 2m\pi$, where m is an integer. But since φ and φ' both lie between $-\pi$ and π, we have $|\varphi - \varphi'| < 2\pi$, so that $m = 0$ and $\varphi = \varphi'$. By (39), the sector (42) is mapped on a region in the z-plane bounded by two circular arcs symmetrical to the real axis and intersecting at $z = 1$ and $z = -1$ at an interior angle of $2\alpha\pi$.

The half plane $\Re w < 1$ is represented in (41) by $-\pi/2 < \varphi < \pi/2$, and consequently mapped in the z-plane on a region similar to the preceding one except that the interior angle is now $\alpha\pi \leqq \pi$, so that this region is interior to $|z|<1$. Since $|w|<1$ is interior to $\Re w<1$, $|w|<1$ is mapped on a simple region D_k interior to $|z|<1$, and it is evident from what precedes that $f(w)$ is holomorphic on $|w| = 1$ except at $w = 1$, and the inverse function holomorphic on the boundary of D_k except at $z = 1$.

To prove that D_{k_1} is interior to D_k when $k_1 > k$, we note that by (39) and (40), (32) becomes

$$w = 1 - (1 - w_1)^{\frac{\alpha_1}{\alpha}},$$

and our proof reduces to showing that this maps $|w_1| < 1$ on a region interior to $|w| < 1$. The polar equation of $|w_1| = 1$ is

$$1 - w_1 = 2\cos\varphi \cdot e^{\varphi i}, \qquad -\pi/2 \leqq \varphi \leqq \pi/2,$$

and writing

$$\beta = \frac{\alpha_1}{\alpha} = 2^{k-k_1} < 1,$$

we therefore have to show that

$$w = 1 - (2\cos\varphi)^{\beta} e^{\beta\varphi i}$$

satisfies $w\bar{w} - 1 < 0$ for $0 \leqq \varphi < \pi/2$ (and hence by symmetry for $-\pi/2 < \varphi \leqq 0$; for $\varphi = \pi/2$ we have $w = 1$). Now write

$$F(\varphi) = \frac{w\bar{w} - 1}{2(2\cos\varphi)^{\beta}} = \frac{1}{2}(2\cos\varphi)^{\beta} - \cos\beta\varphi, \tag{43}$$

whence

$$\frac{\cos\varphi}{\beta}\,\frac{dF(\varphi)}{d\varphi} = -\sin(1-\beta)\varphi - F(\varphi)\sin\varphi. \tag{44}$$

Since $\beta < 1$, $F(0) < 0$ by (43), and therefore at the smallest value of φ between 0 and $\pi/2$ (excl.) for which $F(\varphi) = 0$, we must have $dF/d\varphi \geqq 0$ whence $\sin(1-\beta)\varphi \leqq 0$ by (44), and this is impossible. Hence $F(\varphi) < 0$ for $0 \leqq \varphi < \pi/2$, and D_{k_1} is interior to D_k when $k_1 > k$.

By (38), $\lambda_1 > \lambda > 1$ when $k_1 > k > 0$, and consequently all the conditions for the applicability of Theorems 1 to 3 are satisfied. We may therefore state

THEOREM 4. *Any series summable* (Vk), $k \geqq 0$, *is also summable Poisson to the same sum. A series summable Cesàro of any order is also summable* (Vk) *for any* $k > 0$ *to the same sum. A series summable* (Vk) *is also summable* (Vk_1) *to the same sum when* $k_1 > k \geqq 0$.

In order to investigate the (Vk) summability of the geometric series, it is necessary to consider the boundary curve L_k of the region D_k. A parametric representation of L_k is obtained by introducing the polar equation of $|w| = 1$ in (37):

$$\frac{1-z}{1+z} = (2\cos\varphi)^{\alpha} e^{\alpha\varphi i}, \qquad -\frac{\pi}{2} \leqq \varphi \leqq \frac{\pi}{2}. \tag{L_k}$$

By L_k^{-1} we denote the curve obtained by inversion of L_k with respect to the unit circle, or what is the same on account of L_k being symmetrical about the real axis, by replacing z by $1/z$:

$$(L_k^{-1}) \qquad \frac{z-1}{z+1} = (2\cos\varphi)^\alpha e^{\alpha\varphi i}, \qquad -\frac{\pi}{2} \leqq \varphi \leqq \frac{\pi}{2}.$$

We shall now prove

THEOREM 5. *The geometric series* $1+\zeta+\zeta^2+\cdots$ *is summable* (Vk) *to the sum* $1/(1-\zeta)$ *when* ζ *is inside* L_k^{-1}, *and the summability is uniform for* ζ *in any closed region interior to* L_k^{-1}. *For* ζ *on or outside* L_k^{-1} *the summation is divergent.*

First we dispose of the case $\zeta = 1$; here $u_\nu = 1$ in (9) so that the left hand member is

$$\frac{1}{1-z} = \frac{1}{2} + \frac{1}{2(1-w)^\alpha}$$

by (37). Consequently (9) and (13) give

$$\sum_{n=0}^{\infty} b_n V_n^{(k)} w^n = \frac{1}{2(1-w)^\alpha} + \frac{1}{2(1-w)^{2\alpha}};$$

expanding to the right by (17), comparing coefficients and replacing b_n by its value (36), we find

$$V_n^{(k)} = \frac{1}{2} + \frac{1}{2} \cdot \frac{\Gamma(\alpha)\Gamma(n+2\alpha)}{\Gamma(2\alpha)\Gamma(n+\alpha)} \to \infty \text{ as } n \to \infty.$$

When $\zeta \neq 1$, we proceed as usual by subtracting the sum s from the original u_0, i. e. we now make $u_0 = 1 - 1/(1-\zeta)$, $u_\nu = \zeta^\nu$, $(\nu > 0)$. Writing

$$(45) \qquad z_0 = \frac{1}{\zeta},$$

we have for $|z| < |z_0|$

$$\sum_{\nu=0}^{\infty} u_\nu z^\nu = \frac{1}{1-\zeta z} - \frac{1}{1-\zeta} = -\frac{\zeta}{1-\zeta} \cdot \frac{1-z}{1-\zeta z}$$
$$= \frac{1}{1-\zeta} \cdot \frac{1-z}{z-z_0} = \frac{(1-w)^\alpha}{1-\zeta} \cdot \frac{1+z}{z-z_0},$$

the last step by (37). Using Cauchy's integral formula on (9), we now find

$$(46) \qquad b_n V_n^{(k)} = \frac{1}{2\pi i} \int \frac{1}{1-\zeta} \frac{1+z}{z-z_0} \frac{dw}{w^{n+1}},$$

the integral being taken over a sufficiently small circle around the origin. We shall deform this path of integration into the curve C composed of

the straight line from $w=1$ to $w=1+h$, $(h>0)$, the circle $|w|=1+h$ and the straight line from $w=1+h$ to $w=1$; this deformation is permissible when h is chosen properly. First let ζ belong to a closed region interior to L_k^{-1}, then the z_0 corresponding to any such ζ by (45) is exterior to L_k, and the smallest distance of any such z_0 from L_k is positive, say 2δ. The region interior to C is mapped on a region D in the z-plane, and since $|w|=1$ is interior to C (except that $w=1$ lies on C), L_k is interior to D, except that $z=1$ is on the boundary of D. As $h\to 0$, the boundary of D approaches L_k uniformly along its entire length, and we may therefore choose h so small that every z_0 under consideration is exterior to D, and the smallest distance of any such z_0 to the boundary of D exceeds δ. Consequently, there exists an M such that

$$\left|\frac{1}{1-\zeta}\,\frac{1+z}{z-z_0}\right|<M \tag{47}$$

for ζ in the region considered and z on the boundary of D. Since z_0 is exterior to D, we may replace (46) by

$$\frac{1}{2\pi i}\int_C \frac{1}{1-\zeta}\,\frac{1+z}{z-z_0}\,\frac{dw}{w^{n+1}} = b_n V_n^{(k)}. \tag{48}$$

The integral over each rectilinear part of C is less in absolute value than

$$M\int_1^{1+h}\frac{dw}{w^{n+1}}<M\int_1^{\infty}\frac{dw}{w^{n+1}}=\frac{M}{n},$$

and the absolute value of the integral over the circle $|w|=1+h$ is less than $2\pi M(1+h)^{-n}$. Hence we have

$$\frac{1}{2\pi i}\int_C \frac{1}{1-\zeta}\,\frac{1+z}{z-z_0}\,\frac{dw}{w^{n+1}} = O\left(\frac{1}{n}\right) \tag{49}$$

uniformly for ζ in the region considered, and from (16), (48) and (49) it follows that

$$V_n^{(k)} = O(n^{-\alpha})$$

uniformly in ζ, which proves the first part of the theorem. Now let $\zeta \neq 1$ be on or outside L_k^{-1}; then, by (45), $z_0 \neq 1$ is on or inside L_k (and hence interior to D, no restriction on h being necessary here). The point $w_0 \neq 1$ corresponding to z_0 is therefore interior to or on the unit circle, and (48) is replaced by

$$\frac{1}{2\pi i}\int_C \frac{1}{1-\zeta}\,\frac{1+z}{z-z_0}\,\frac{dw}{w^{n+1}} = b_n V_n^{(k)} + \text{residue at } w_0, \tag{50}$$

the residue being equal to

$$\frac{1}{1-\zeta}\,\frac{1+z_0}{(dz/dw)_0}\cdot\frac{1}{w_0^{n+1}} = \frac{2(1-w_0)}{\alpha(1-\zeta)(1-z_0)}\cdot\frac{1}{w_0^{n+1}}. \tag{51}$$

Evidently (49) holds for the particular ζ under consideration and since $|w_0| \leqq 1$, $w_0 \neq 1$, (50) and (51) show that $V_n^{(k)}$ does not approach a finite limit as $n \to \infty$, and this completes the proof of Theorem 5.

It is appropriate to note here that any ζ not on the real axis from 1 to $+\infty$ is interior to an L_k^{-1} provided that k is chosen large enough. In fact, this is equivalent to stating that any $z = 1/\zeta$ not real and between 0 and 1 inclusive, is exterior to L_k, or by (39), that any Z not real and between the same limits is exterior to the Z-transform of L_k which is

$$Z = (2\cos\varphi)^\alpha e^{\alpha\varphi i}, \quad -\pi/2 \leqq \varphi \leqq \pi/2. \tag{52}$$

Since the abscissa of this curve lies between 0 and 2^α, inclusive, and the absolute value of the ordinate does not exceed $2^\alpha \sin(\alpha\pi/2)$, the curve approaches the line segment $0 \leqq Z \leqq 1$ as $\alpha \to 0$, i. e. $k \to \infty$, and the statement follows.

In concluding this paragraph, we shall prove the following recurrent formula connecting (Vk) with $(V, k+1)$:

$$b_n^{(k+1)}\, V_n^{(k+1)} = \sum_{\nu=1}^{n} \frac{\nu(2n-\nu-1)!}{2^{2n-\nu}(n-\nu)!\; n!}\, b_\nu^{(k)}\, V_\nu^{(k)} \tag{53}$$

for $n \geqq 1$. (When $n = 0$, we have the trivial formula $V_0^{(k+1)} = V_0^{(k)} = u_0$). Denoting by α_1 and w_1 the α and w belonging to $(V, k+1)$, we have, by (13), $\alpha_1 = 2^{-k-1} = \alpha/2$, and by (37)

$$\frac{1-z}{1+z} = (1-w)^\alpha = (1-w_1)^{\alpha/2},$$

so that

$$1-w = (1-w_1)^{1/2}; \tag{54}$$

moreover, (13) gives $g(w) = g_1(w_1)$. From (9) it therefore follows that

$$\sum_{\nu=0}^{\infty} b_\nu^{(k)}\, V_\nu^{(k)}\, w^\nu = \sum_{n=0}^{\infty} b_n^{(k+1)}\, V_n^{(k+1)}\, w_1^n,$$

and (53) is obtained by substitution of the power series for w^ν to the left and comparison of coefficients of w_1^n. All we have to show is consequently that

$$w^\nu = \sum_{n=\nu}^{\infty} \frac{\nu(2n-\nu-1)!}{2^{2n-\nu}(n-\nu)!\; n!}\, w_1^n, \tag{55}$$

where for $\nu = 0$, the constant term to the right is to be replaced by unity. For $\nu = 0$, (55) is evident, and for $\nu = 1$, it is obtained by applying the binomial theorem to (54). For $\nu > 1$, we observe that (54) may be written $w^2 - 2w + w_1 = 0$, whence the recurrent formula

$$w^{\nu+1} = 2w^\nu - w_1 w^{\nu-1},$$

and the verification of (55) proceeds in the same way as in paragraph 2.

7. **The generalized Euler summation (Ep) of order $p \geqq 0$.** The nth (Ep) sum we denote by $E_n^{(p)}$, and in order to obtain formal agreement with the work of Knopp,[8] we define it not as U_n but as $b_n U_n$ in (9) when $f(w)$ and $g(w)$ are given by (14). This deviation from our general definition is irrelevant, since by (14)

$$g(w) = \frac{1}{1-w} - \frac{1-\vartheta}{1-(1-\vartheta)w},$$

and consequently

$$b_n = 1-(1-\vartheta)^{n+1},$$

so that $b_n \to 1$ as $n \to \infty$.

Written out, our definition of $E_n^{(p)}$ is

$$\sum_{\nu=0}^{\infty} u_\nu z^\nu = \frac{(1-w)(1-(1-\vartheta)w)}{\vartheta} \sum_{n=0}^{\infty} E_n^{(p)} w^n, \tag{56}$$

where $\vartheta = 2^{-p}$ and

$$w = \frac{z}{\vartheta + (1-\vartheta)z}, \quad z = \frac{\vartheta w}{1-(1-\vartheta)w}, \quad \frac{z}{1-z} = \vartheta \cdot \frac{w}{1-w}. \tag{57}$$

Comparison of (5) and (57) gives $\lambda = 1$, $a = 1/\vartheta$, and the map of $|w| < 1$ is the circle

$$(2-\vartheta)z\bar{z} - (1-\vartheta)(z+\bar{z}) - \vartheta < 0, \tag{58}$$

or

$$\left| z - \frac{1-\vartheta}{2-\vartheta} \right| < \frac{1}{2-\vartheta}, \tag{58a}$$

which is evidently interior to $|z| < 1$.

We shall now show that our definition of $E_n^{(p)}$ coincides with that of Knopp, and also that any series summable (Ep) is also summable (Ep_1) to the same sum when $p_1 > p \geqq 0$.[10] Both these statements are proved by setting up the relation between (Ep_1) and (Ep). Eliminating z between (57) and the corresponding equation where ϑ and w are replaced by ϑ_1 and w_1, we find

[10] Knopp, l. c. [8] p. 247 and following. Our Theorem 3 is inapplicable here, since $\lambda_1 = \lambda = 1$ (although D_1 is interior to D).

$$w = \frac{h w_1}{1-(1-h)w_1}, \tag{59}$$

where

$$h = \frac{\vartheta_1}{\vartheta}. \tag{60}$$

From (14) and (59)

$$\frac{g_1(w_1)}{g(w)} = \frac{\vartheta_1(1-w)(1-(1-\vartheta)w)}{\vartheta(1-w_1)(1-(1-\vartheta_1)w_1)} = \frac{h}{(1-(1-h)w_1)^2},$$

and (56) gives

$$\sum_{n=0}^{\infty} E_n^{(p_1)} w_1^n = \frac{h}{(1-(1-h)w_1)^2} \sum_{\nu=0}^{\infty} E_\nu^{(p)} w^\nu$$
$$= \sum_{\nu=0}^{\infty} E_\nu^{(p)} \frac{h^{\nu+1} w_1^\nu}{(1-(1-h)w_1)^{\nu+2}}.$$

By the binomial theorem

$$\frac{w_1^\nu}{(1-(1-h)w_1)^{\nu+2}} = \sum_{n=\nu}^{\infty} \binom{n+1}{\nu+1} (1-h)^{n-\nu} w_1^n,$$

and introducing this in the preceding equation, we find

$$E_n^{(p_1)} = \sum_{\nu=0}^{n} \binom{n+1}{\nu+1} h^{\nu+1} (1-h)^{n-\nu} E_\nu^{(p)}. \tag{61}$$

When $p = 0$, $\vartheta = 1$, $w = z$, and (56) becomes

$$\sum_{\nu=0}^{\infty} u_\nu z^\nu = (1-z) \sum_{n=0}^{\infty} E_n^{(0)} z^n,$$

whence

$$E_n^{(0)} = u_0 + u_1 + \cdots + u_n. \tag{62}$$

In (61) we now make $p = 0$, so that $h = \vartheta_1$; replacing p_1 and ϑ_1 by p and ϑ, we therefore have with the aid of (62)

$$E_n^{(p)} = \sum_{\nu=0}^{n} \binom{n+1}{\nu+1} \vartheta^{\nu+1} (1-\vartheta)^{n-\nu} (u_0 + u_1 + \cdots + u_\nu). \tag{63}$$

Except for changes in notation, this is Knopp's definition.[11]

Now let $u_0 + u_1 + u_2 + \cdots$ be summable (Ep) to the sum zero, so that, for any $\varepsilon > 0$, $|E_\nu^{(p)}| < \varepsilon$ for $\nu \geq m = m(\varepsilon)$. When $p_1 > p$, we have $0 < h < 1$ by (60), and each of the m first terms in (61) contains the factor $(1-h)^{n-m}$, so that the sum of these terms approaches zero as $n \to \infty$. The sum of the remaining terms in (61) is less in absolute value than

[11] L. c. [8], p. 245.

$$\varepsilon \sum_{\nu=m}^{n} \binom{n+1}{\nu+1} h^{\nu+1}(1-h)^{n-\nu}$$

$$= \varepsilon \sum_{\nu=m+1}^{n+1} \binom{n+1}{\nu} h^{\nu}(1-h)^{n+1-\nu} < \varepsilon (h+1-h)^{n+1} = \varepsilon,$$

and consequently $E_n^{(p_1)} \to 0$ as $n \to \infty$.

We shall finally prove the following theorem, which establishes the connection between the summation methods (Ep) and (Vk):

THEOREM 6. *A series which is summable* (Ep), $p \geqq 0$, *is also summable* (Vk), $k > 0$, *to the same sum when*

$$\frac{1}{2^{\alpha}} + \vartheta > 1, \qquad (\alpha = 2^{-k}, \quad \vartheta = 2^{-p}). \tag{64}$$

Since we assumed $k > 0$, the λ in (Vk) is > 1, while that in (Ep) is $= 1$, and our theorem follows from Theorem 3 as soon as we have shown that L_k is interior to the circle (58) (except that $z = 1$ is common to both boundaries). By (39) the map of (58) on the Z-plane is the circle

$$2(1-\vartheta)\, Z\bar{Z} - Z - \bar{Z} < 0,$$

and all we have to show is that (52), the Z-map of L_k, is interior to this circle (the point $Z = 0$ excepted). In other words, we must show that $F(\varphi) < 0$ for $-\pi/2 < \varphi < \pi/2$, where

$$F(\varphi) = (1-\vartheta)\,(2\cos\varphi)^{\alpha} - \cos\alpha\varphi.$$

Now (64) is identical with $F(0) < 0$, and we have

$$\frac{\cos\varphi}{\alpha}\,\frac{dF(\varphi)}{d\varphi} = -\sin(1-\alpha)\varphi - F(\varphi)\sin\varphi;$$

except for the notation, this is the same as (44), and the proof proceeds as there.

DEPARTMENT OF PHYSICS,
COLUMBIA UNIVERSITY.

A RULE OF SIGNS INVOLVING CERTAIN ORTHOGONAL POLYNOMIALS.[1]

By Morris Marden.

At the beginning of last summer, in a lecture delivered to Mathematical Colloquium at Zurich, Switzerland, Professor Shohat raised the question as to whether a rule of signs involving Legendre polynomials, as given by Laguerre in his Oeuvres, t. I, pp. 144–146, could be carried over to other sets of orthogonal polynomials.[2] The present note attempts to answer this question, following the general lines of reasoning given by Laguerre. It suggests a rule of signs involving a set of orthogonal polynomials $P_n(x)$ which may be defined as follows.

1. **The polynomials $P_n(x)$.**

(a) *Definition.* The polynomials $P_n(x)$ are to have the form

$$P_n(x) = x^n + c_1 x^{n-1} + \cdots + c_n, \qquad (n = 0, 1, 2, \cdots)$$

and satisfy the orthogonality conditions

$$\int_a^b p(x)\, P_m(x)\, P_n(x)\, dx = 0, \qquad m \neq n. \tag{1.11}$$

The weight-function $p(x)$ is to be non-negative on the interval $-\infty \leqq a \leqq x \leqq b \leqq \infty$, and satisfy the equation

$$\frac{d}{dx}[K(x)\, p(x)] = L(x)\, p(x) \tag{1.12}$$

where

$$\begin{aligned} K(x) &= Ax^2 + Bx + C, \\ L(x) &= Dx + E, \end{aligned} \tag{1.13}$$

$$AD \geqq 0, \tag{1.14}$$

and

$$\lim_{x \to a} K(x)\, p(x) = \lim_{x \to b} K(x)\, p(x) = 0. \tag{1.15}$$

Below are tabulated some well-known sets of orthogonal polynomials which are special cases of these polynomials $P_n(x)$.

[1] Received November 15, 1930.—Presented to the Society on September 12, 1930.

[2] The author wishes to thank Professor Pólya for reference to Laguerre's article.

Polynomials	$p(x)$	A	B	C	D	E	a	b
Legendre	1	-1	0	1	-2	0	-1	1
Tschebyscheff	$(1-x^2)^{\pm 1/2}$	-1	0	1	$-(2 \pm 1)$	0	-1	1
Jacobi	$(1-x)^\alpha (1+x)^\beta$	-1	0	1	$-(\alpha+\beta+2) \leqq 0$	$\beta-\alpha$	-1	1
Laguerre	$e^{-x} x^\alpha$	0	1	0	-1	$\alpha+1$	0	∞
Hermite	$e^{-x^2/2}$	0	0	1	-1	0	$-\infty$	∞

Among the polynomials $P_n(x)$ there are, however, none which differ "essentially" from the five kinds listed above. For, equation (1.12) has the following integrals:

1. If $A \neq 0$ and $q = B^2 - 4AC < 0$,

$$p(x)\,K(x) = [K(x)]^{D/A} \exp\left(\frac{2F}{\sqrt{-q}} \tan^{-1} \frac{2Ax+B}{\sqrt{-q}}\right)$$

where $F = \frac{1}{2A}(2AE - BD)$.

2. If $A \neq 0$ and $q > 0$,

$$p(x)\,K(x) = \lambda (x - x_1)^\mu (x - x_2)^\nu$$

where $\lambda = A^{D/A}$, $\mu = \frac{D}{A} + \frac{F}{\sqrt{q}}$ and $\nu = \frac{D}{A} - \frac{F}{\sqrt{q}}$, and $K(x_1) = K(x_2) = 0$.

3. If $A = 0$, but $B \neq 0$,

$$p(x)\,K(x) = e^{(D/B)x} (Bx + C)^G$$

where $G = \frac{1}{B^2}(BE - CD)$.

4. Finally, if $A = B = 0$,

$$p(x)\,K(x) = \exp.\left(\frac{D}{2C} x^2 + \frac{Ex}{C}\right).$$

If the first integral is to fulfil requirement (1.15), $K(a) = K(b) = 0$, which is impossible. For, its discriminant q being negative, $K(x)$ cannot have any real roots. If the second integral is to fulfil requirement (1.15), then

$$a = x_1, \qquad b = x_2, \qquad \mu > 0, \qquad \nu < 0.$$

For the third integral to satisfy condition (1.15), then $G > 0$ and eithe $D/B > 0$ in which case $a = -\infty$, $K(b) = 0$, or $D/B < 0$ in which case $K(a) = 0$ and $b = +\infty$. Finally, for the fourth integral to satisfy condition (1.15)

$$D/C < 0 \quad \text{and} \quad a = -\infty, \quad b = +\infty.$$

In short, the polynomials $P_n(x)$ include only the Jacobi, Laguerre and Hermite polynomials and such polynomials as can be reduced to one of these three kinds by affine transformations.

(b) *Properties.* As orthogonal polynomials[3] *the* $P_n(x)$ *have only real and simple zeros, all lying in the interval* $a<x<b$.

These polynomials furthermore satisfy the differential equation

$$K(x)\,P_n''(x)+L(x)\,P_n'(x)-n[A(n-1)+D]\,P_n(x)=0. \tag{1.21}$$

For,[4] by twice integrating by parts and using condition (1.15), one finds

$$\begin{aligned}\int_a^b x^m \frac{d}{dx}[K(x)\,p(x)\,P_n(x)]\,dx &= -m\int_a^b x^{m-1}K(x)\,p(x)\,P_n'(x)\,dx \\ &= m\int_a^b P_n(x)\frac{d}{dx}[x^{m-1}K(x)\,p(x)]\,dx.\end{aligned} \tag{1.22}$$

Now

$$\frac{d}{dx}[x^{m-1}K(x)\,p(x)] = [(m-1)\,x^{m-2}K(x)+x^{m-1}L(x)]\,p(x), \tag{1.23}$$

where the brackets contain a polynomial of at most the mth degree. Hence by condition (1.11) the right hand side of (1.22) vanishes for $m\leqq n-1$. This means, since

$$\frac{d}{dx}[K(x)\,p(x)\,P_n'(x)] = [K(x)\,P_n''(x)+L(x)\,P_n'(x)]\,p(x), \tag{1.24}$$

that the polynomial of nth degree contained in the bracket on the right hand side of (1.24) differs from $P_n(x)$ only by a multiplicative constant. That is to say,

$$K(x)\,P_n''(x)+L(x)\,P_n'(x) = M(n)\,P_n(x). \tag{1.25}$$

By comparison of the two sides of equation (1.25) one calculates $M(n)$ to be

$$M(n) = n[A(n-1)+D]. \tag{1.26}$$

2. **The rule of signs.** In what follows it will be assumed that b is a finite number. This restricts the polynomials $P_n(x)$ to be essentially (cf. § 1 (a)) either Jacobi or Laguerre polynomials.

Our principal theorem is then the following:

Given any real polynomial $F(x)$ *of degree* n. *By suitable choice of constants* λ_j, $F(x)$ *may be written as*

$$F(x) = \lambda_0 P_0(x)+\lambda_1 P_1(x)+\cdots+\lambda_n P_n(x).$$

[3] Stieltjes Œuvres, vol. I, p. 380.

[4] Cf. Pólya-Szegö, Aufgaben und Lehrsätze, Berlin (1925), vol. II, pp. 75–97.

Then, the number of real zeros of $F(x)$ greater or equal to b does not exceed the number of variations of sign in the sequence $\lambda_0, \lambda_1, \cdots, \lambda_n$.

Throughout this note the symbol ${}_0\varphi$ will be used to denote the number of real zeros which a given function $\varphi(x)$ has in the interval $x \geqq b$. The quantity $\operatorname{sg} x$ will be taken as equal to $+1$, -1, or 0, according as x is positive, negative or zero.

If V represents the number of variations of sign in the sequence $\lambda_0, \lambda_1, \cdots, \lambda_n$, then the relation to be proved is that ${}_0F \leqq V$.

The theorem is evidently true when $V = 0$. For, as $P_n(x) \neq 0$ when $x \geqq b$,

$$\operatorname{sg} P_n(x) = \operatorname{sg} P_n(\infty) = 1, \qquad x \geqq b.$$

The constants λ_j all being of like sign, the function $F(x)$ cannot vanish anywhere in the interval $x \geqq b$; that is, ${}_0F = 0$.

Suppose the theorem were true for $V-1$ variations of sign in the sequence $\lambda_0, \lambda_1, \cdots, \lambda_n$. Would it therefore be true when there are V variations of sign?

To answer this question, we follow the reasoning of Laguerre (l. c.). Assume that

$$\lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_{k+l-1} = 0, \quad \text{but} \quad \lambda_k \lambda_{k+l} < 0.$$

As, in any interval $b \leqq x \leqq c < \infty$, the function $R(x) = \dfrac{F(x)}{P_{k+l}(x)}$ has a derivative $R'(x) = \dfrac{f(x)}{[P_{k+l}(x)]^2}$ where

$$f(x) = P_{k+l}(x)\, F'(x) - P'_{k+l}(x)\, F(x), \tag{2.2}$$

it follows from Rolle's theorem that

$${}_0F \leqq {}_0f + 1. \tag{2.3}$$

Now, from relation (2.2) one deduces that

$$f(x) = \sum_{j=0}^{n} \lambda_j \begin{vmatrix} P'_j(x) & P'_{k+l}(x) \\ P_j(x) & P_{k+l}(x) \end{vmatrix}, \tag{2.4}$$

and from relation (1.21) that

$$\begin{aligned} &\frac{d}{dx}\left\{K(x) \begin{vmatrix} P'_j(x) & P'_{k+l}(x) \\ P_j(x) & P_{k+l}(x) \end{vmatrix} p(x)\right\} \\ &= [M(j) - M(k+l)]\, p(x)\, P_j(x)\, P_{k+l}(x) \end{aligned} \tag{2.5}$$

where

$$\begin{aligned} M(j) - M(k+l) &= j[A(j-1)+D] - (k+l)[A(k+l-1)+D] \\ &= (j-k-l)[A(j+k+l-1)+D]. \end{aligned} \tag{2.6}$$

Because of condition (1.14)

$$\operatorname{sg}[M(j)-M(k+l)] = \operatorname{sg}(j-k-l)\operatorname{sg}(A+D).$$

On combining equations (2.4) and (2.5), one obtains

$$\frac{d}{dx}\left\{K(x)p(x)f(x)\right\} = p(x)\Phi(x)P_{k+l}(x) \tag{2.7}$$

where

$$\Phi(x) = \sum_{j=0}^{n}\lambda_j[M(j)-M(k+l)]P_j(x).$$

Now, let $\pi(x) = K(x)p(x)f(x)$ and $\pi'(x) = p(x)\Phi(x)P_{k+l}(x)$. Then ${}_0\pi = {}_0K+{}_0p+{}_0f$, ${}_0\pi' = {}_0p+{}_0\Phi$. By Rolle's theorem, ${}_0\pi \leqq {}_0\pi'+1$; i. e. ${}_0f \leqq {}_0\Phi+1-{}_0K$. As ${}_0K = 1$ (cf. § 1(a)),

$${}_0f \leqq {}_0\Phi. \tag{2.8}$$

On the other hand, $\Phi(x)$ is a linear combination of the $P_j(x)$ with only $V-1$ changes of sign in its coefficients.

By hypothesis, therefore,

$${}_0\Phi \leqq V-1. \tag{2.9}$$

On combining results (2.3), (2.8) and (2.9), one finally concludes that ${}_0F \leqq V$, as was to be proved.

This proof and, consequently, the rule of signs are equally valid, in the interval $b \leqq x \leqq c \leqq \infty$, for an infinite expansion

$$F(x) = \lambda_0 P_0(x)+\lambda_1 P_1(x)+\lambda_2 P_2(x)+\cdots,$$

where the sequence of coefficients $\lambda_0, \lambda_1, \lambda_2, \cdots$ shows but a finite number of variations of sign, provided that each of the two series

$$(1)\quad \sum_{j=1}^{\infty} j\lambda_j[A(j-1)+D]P_j(x), \qquad (2)\quad \sum_{j=1}^{\infty}\lambda_j P_j'(x),$$

converges uniformly in the interval $b \leqq x \leqq c$. For, the summations and differentiations performed in the proof are legitimate when the three series

$$(4)\quad \sum_{j=0}^{\infty}\lambda_j P_j(x), \qquad (5)\quad \sum_{j=1}^{\infty}\lambda_j P_j'(x), \qquad (6)\quad \sum_{j=2}^{\infty}\lambda_j P_j''(x),$$

converge uniformly. Now, the uniform convergence of series (4) is guaranteed by that of series (1), and, due to equation (1.21), the uniform convergence of series (6) is assured by that of the series (1) and (2).

3. **Corollaries.** The rule of signs of the preceding paragraph differs from Descartes' rule only in two respects. First, the polynomial $P_k(x)$ replaces the polynomial x^k of Descartes' rule and secondly, the interval (a, b) wherein lie all the zeros of the polynomials $P_k(x)$ replaces the point $x = 0$ of Descartes' rule.

This analogy is carried further in the corollaries which will now be stated.

COROLLARY 1. *If there is a lacuna of length $k+1$ in the sequence $\lambda_0, \lambda_1, \cdots, \lambda_n$, then $f(x)$ has at least k zeros which are not on the real axis in the interval $x \geqq b$. (That is to say, there are k zeros which are either imaginary or real and less than b.)*

COROLLARY 2. *If $p(a+b-x) = p(x)$, then the number of real zeros of $F(x)$ less than or equal to a does not exceed the number of variations of sign in the sequence $\lambda_0, -\lambda_1, \lambda_2, \cdots, (-1)^n \lambda_n$.*

COROLLARY 3. *If $p(a+b-x) = p(x)$ and if all the zeros of $F(x)$ lie on the real axis in the intervals $x \leqq a$, $x \geqq b$, then the number of zeros greater than or equal to b is exactly equal to the number of variations of sign in the sequence $\lambda_0, \lambda_1, \cdots, \lambda_n$, and the number of zeros less than or equal to a is exactly equal to the number of variations of sign in the sequence*

$$\lambda_0, -\lambda_1, \lambda_2, \cdots, (-1)^n \lambda_n.$$

To prove the first corollary, suppose that

$$\lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_{r+k} = 0 \quad \text{but} \quad \lambda_r \lambda_{r+k+1} \neq 0.$$

Then $V \leqq n-k$ and hence ${}_0F \leqq n-k$ or $n-{}_0F \geqq k$.

To prove the second corollary note that, because of the requirement $p(a+b-x) = p(x)$ (which requirement restricts the $P_n(x)$ to be Jacobi polynomials with $\alpha = \beta$),

$$\int_a^b p(x)\, P_n(x)\, (a+b-x)^m\, dx = -\int_a^b p(a+b-x)\, P_n(a+b-x) x^m\, dx$$
$$= -\int_a^b p(x)\, P_n(a+b-x) x^m\, dx.$$

The left hand side of this equation being zero for $m \leqq n-1$, so is the lower right hand side. Hence, $P_n(a+b-x)$ differs from $P_n(x)$ only by a multiplicative constant. That is $P_n(a+b-x) = (-1)^n P_n(x)$. Accordingly,

$$F(a+b-x) = \lambda_0 P_0(x) - \lambda_1 P_1(x) + \lambda_2 P_2(x) + \cdots + (-1)^n \lambda_n P_n(x).$$

Consequently, by Corollary 1, the number of zeros for $a+b-x \geqq b$, i. e. $x \leqq a$, does not exceed the number of variations of sign in the sequence $\lambda_0, -\lambda_1, \lambda_2, \cdots, (-1)^n \lambda_n$.

To establish the last corollary, let ${}_0F'$ denote the number of real zeros of F which are equal to or less than a, and V' the number of variations of sign in the sequence

$$\lambda_0, -\lambda_1, \lambda_2, \cdots, (-1)^n \lambda_n.$$

Then ${}_0F + {}_0F' = n$. Since $V + V' \leqq n$ and ${}_0F \leqq V$, ${}_0F' \leqq V'$, it follows that ${}_0F = V$ and ${}_0F' = V'$.

The writer is now considering the possibility of a similar rule of signs involving arbitrary orthogonal polynomials.

UNIVERSITY OF WISCONSIN AT MILWAUKEE.

ON A CERTAIN INEQUALITY OF STEINER.[1]

By E. J. McShane.[2]

One of the most useful inequalities in problems dealing with the areas of surfaces is that of Steiner.[3] The inequality of Steiner can be stated thus: *Given a region D of the xy plane, and two surfaces $z = f_1(x, y)$, $z = f_2(x, y)$ ((x, y) on D). Then the area of the surface*

$$z = \tfrac{1}{2}(f_1(x, y) + f_2(x, y))$$

is equal to or less than the arithmetic mean of the areas of the surfaces $z = f_1$, $z = f_2$. In enunciating this theorem Steiner did not mention the definition of area which he used, nor did he state the restrictions upon the functions f_1, f_2. It may easily be shown that if the area be defined in the sense of Lebesgue, then the inequality is valid for all continuous surfaces $z = f_1(x, y)$, $z = f_2(x, y)$.

For some purposes it is desirable to exclude the sign of equality from this relation. Section 2 of this paper is devoted to finding conditions under which the inequality holds in the strict sense, the remaining sections to certain applications of the inequality. The first such application is the finding of a necessary condition that the areas of a sequence of polyhedra tending to a surface $z = z(x, y)$ shall tend to the area of that surface. If we examine the well-known example of Schwarz, of a sequence of polyhedra inscribed in a circular cylinder, having the maximum length of side tending to zero and the areas increasing without limit, we see that this necessary condition is not fulfilled.

The second application is that of establishing the analytic character of the surface of least area. The results are contained in Theorems VII to X. Since these theorems are of interest in themselves, quite apart from the preceding sections, a second proof (of an elementary nature) is given for Theorem VII, which makes it possible to read from § 4 on without reference to preceding demonstrations.

I wish to thank Prof. Radó, first for directing my attention to this problem, and second for a suggestion which enabled me to replace certain less elegant results by the present §§ 2, 3.

1. **Preliminaries.** In order to avoid repetition, we shall understand throughout that each rectangle referred to is assumed to have its sides

[1] Received March, 1931.

[2] National Research Fellow.

[3] J. Steiner, Gesammelte Werke, Bd. II, S. 298.

parallel to the coördinate axes: $a_1 \leqq x \leqq b_1$, $a_2 \leqq y \leqq b_2$. The ordinary notation for the partial derivatives of a function $z(x, y)$ will be used: $p = \dfrac{\partial z}{\partial x}$, $q = \dfrac{\partial z}{\partial y}$; and in case several functions are present we shall use the same subscripts on the p and q as on the z:

$$p_n = \frac{\partial z_n}{\partial x}, \quad q_n = \frac{\partial z_n}{\partial y}, \quad \pi = \frac{\partial \zeta}{\partial x}, \quad \varkappa = \frac{\partial \zeta}{\partial y},$$

etc. The frequently occurring radical $\sqrt{1+\left(\dfrac{\partial f}{\partial x}\right)^2+\left(\dfrac{\partial f}{\partial y}\right)^2}$ will be denoted by $w(f; x, y)$, or simply by $w(f)$; and the area of the surface defined by $z = z(x, y)$ ((x, y) on Q) will be denoted by $A_Q(z)$.

2. **The inequality of Steiner.** In the theorems of this section we shall have occasion to use the partial derivatives $p(x, y)$, $q(x, y)$ of a function $z(x, y)$ ((x, y) on Q), knowing only that $A_Q(z)$ is finite. That these derivatives exist almost everywhere in Q is a consequence of the theorem[4] that if on the square Q: $(a_1 \leqq x \leqq b_1, a_2 \leqq y \leqq b_2)$ the surface $z = z(x, y)$ has finite area, then $z(x, y)$ is of limited total variation in x for almost all fixed values of y in the interval (a_2, b_2) and of limited total variation in y for almost all fixed values of x in the interval (a_1, b_1). The functions $p(x, y)$, $q(x, y)$ are measurable.[5]

Let us now turn our attention to surfaces $z = f(x, y)$ ((x, y) on Q), where Q is a square. If we are given two such surfaces $z = z_1(x, y)$, $z = z_2(x, y)$, we can then state that

$$A_Q([z_1+z_2]/2) \leqq \tfrac{1}{2}(A_Q(z_1)+A_Q(z_2)).$$

Our object in this section is to establish conditions under which the equality is excluded; such conditions are established in Theorems I and II. But before proceeding to these theorems we shall give examples to show that the equality cannot always be excluded. A trivial case is that in which $z_1 - z_2$ is constant. More important, we can give an example in which $z_1 - z_2$ is not constant, and yet

$$A_Q([z_1+z_2]/2) = A_Q(z_1) = A_Q(z_2).$$

Let $\varphi(x)$ be a continuous[6] monotonically increasing function such that $\varphi'(x) = 0$ almost everywhere, $\varphi(0) = 0$, $\varphi(1) = 1$. The length of the curve $y = k\varphi(x)$ $(0 \leqq x \leqq 1)$ is easily shown to be equal to $1+|k|$. Define now

[4] L. Tonelli, *Sulla quadratura delle superficie*, Rend. Lincei, ser. VI, vol. III, p. 357. T. Radó, *Sur l'aire des surfaces courbes*. Acta Szeged, vol. 3 (1927), p. 165.

[5] Carathéodory, *Vorlesungen über reelle Funktionen*, p. 642.

[6] For an example of such a function, see, e. g., Carathéodory, Vorlesungen über reelle Funktionen, § 509.

$$z_1(x, y) = 0 \qquad (0 \leqq x \leqq 1,\ 0 \leqq y \leqq 2),$$
$$z_1(x, y) = \varphi(x-1) \qquad (1 \leqq x \leqq 2,\ 0 \leqq y \leqq 2);$$

and define

$$z_2(x, y) = \varphi(x) \qquad (0 \leqq x \leqq 1,\ 0 \leqq y \leqq 2),$$
$$z_2(x, y) = 1 \qquad (1 \leqq x \leqq 2,\ 0 \leqq y \leqq 2).$$

These surfaces are cylinders with generators of length 2 parallel to the y-axis; and we readily calculate that

$$A_Q([z_1+z_2]/2) = A_Q(z_1) = A_Q(z_2) = 6.$$

Having seen that the equality cannot always be excluded, let us establish conditions under which the exclusion is possible. We first restrict ourselves to the particular class of continuous surfaces whose area is given by the ordinary double integral; we can then state:

THEOREM I. *If the functions $z_1(x, y)$, $z_2(x, y)$ are defined and continuous on a square Q, and $A_Q(z_1)$ and $A_Q(z_2)$ are both finite, and*

$$A_Q(z_j) = \int\int_Q w(z_j;\ x, y)\, dx\, dy \qquad (j = 1, 2),$$

then either

$$A_Q([z_1+z_2]/2) < \tfrac{1}{2}(A_Q(z_1)+A_Q(z_2))$$

or $z_1 - z_2$ is constant.

Tonelli has shown[7] that if $A_Q(z)$ is finite, then

$$A_Q(z) = \int\int_Q w(z;\ x, y)\, dx\, dy$$

if and only if $z(x, y)$ is absolutely continuous in y for almost all fixed values of x, and absolutely continuous in x for almost all fixed values of y. Such a function will henceforth be called *absolutely continuous.* (In particular, if $z(x, y)$ represents a polyhedron it is absolutely continuous.) It is clear that if z_1 and z_2 are both absolutely continuous, then so is $(z_1+z_2)/2$. Hence we have

$$\tfrac{1}{2}(A_Q(z_1)+A_Q(z_2)) - A_Q([z_1+z_2]/2)$$
$$= \int\int_Q [\tfrac{1}{2}(w(z_1)+w(z_2)) - w([z_1+z_2]/2)]\, dx\, dy.$$

But the integrand on the right is never negative, and is greater than zero if $w([z_1+z_2]/2)$ is finite and

$$\sqrt{(p_1-p_2)^2+(q_1-q_2)^2} \neq 0;$$

[7] L. Tonelli, loc. cit. [4], p. 633.

for $\frac{1}{2}w(z_1)$, $\frac{1}{2}w(z_2)$ and $w([z_1+z_2]/2)$ are the sides of the triangle whose vertices are $(0,0,0)$, $(p_1/2, q_1/2, 1/2)$ and $([p_1+p_2]/2, [q_1+q_2]/2, 1)$. Hence either the left member of the equation is positive or

$$\sqrt{(p_1-p_2)^2+(q_1-q_2)^2} = 0$$

almost everywhere. We now show that in the latter case z_1-z_2 is constant.

Let the set E consist of

1) all lines $x = x_0$ such that $z_j(x_0, y)$ are not both absolutely continuous in y;
2) all lines $y = y_0$ such that $z_j(x, y_0)$ are not both absolutely continuous in x;
3) all points (x, y) such that $p_1(x, y)$, $p_2(x, y)$, $q_1(x, y)$, and $q_2(x, y)$ are not all defined;
4) all points (x, y) at which

$$\sqrt{(p_1-p_2)^2+(q_1-q_2)^2} \neq 0.$$

This set has measure zero. Hence for almost all points (x_0, y_0) of Q the lines $x = x_0$ and $y = y_0$ have in common with E at most a set of linear measure zero. Fix on one such point (x_0, y_0), and let (x, y) be any other point with the same property. Then

$$z_1(x, y) - z_1(x_0, y_0) = \int_{x_0}^{x} p_1(x, y_0)\,dx + \int_{y_0}^{y} q_1(x, y)\,dy,$$
$$z_2(x, y) - z_2(x_0, y_0) = \int_{x_0}^{x} p_2(x, y_0)\,dx + \int_{y_0}^{y} q_2(x, y)\,dy.$$

But except on a set of linear measure zero the integrands on the right are equal; hence

$$z_1(x, y) - z_2(x, y) = z_1(x_0, y_0) - z_2(x_0, y_0).$$

This being true for almost all (x, y), and z_1 and z_2 being continuous, it remains true for all (x, y) in Q.

In the above demonstration we have made use of the hypothesis that the areas $A_Q(z_1)$ and $A_Q(z_2)$ are given by the classical double integral. Since this imposes an essential restriction upon the class of surfaces considered, it is desirable to develop a theorem in which the surfaces are assumed merely to be continuous. We therefore state

THEOREM II. *If the functions $z_1(x, y)$, $z_2(x, y)$ are defined and continuous on a square Q, and $A_Q(z_1)$ and $A_Q(z_2)$ are both finite, and*

$$\sqrt{(p_1-p_2)^2+(q_1-q_2)^2} > 0$$

on a set of positive measure, then

$$A_Q([z_1+z_2]/2) < \tfrac{1}{2}(A_Q(z_1)+A_Q(z_2)).$$

Let us designate by ζ the function $[z_1(x, y)+z_2(x, y)]/2$. The function of rectangles $\frac{1}{2}(A_R(z_1)+A_R(z_2))-A_R(\zeta)$ is additive and nowhere negative; hence if it be zero when $R = Q$, it must be zero for every R contained in Q, and its derivative must therefore be identically zero. But its derivative is almost everywhere[8] equal to

$$\tfrac{1}{2}(\sqrt{1+p_1^2+q_1^2}+\sqrt{1+p_2^2+q_2^2})-\sqrt{1+\pi^2+\varkappa^2};$$

and on the point set (of positive measure) on which p_1, q_1, p_2, q_2 are finite and $\sqrt{(p_1-p_2)^2+(q_1-q_2)^2}>0$ this is positive. Hence the assumption that

$$\tfrac{1}{2}(A_Q(z_1)+A_Q(z_2))-A_Q(\zeta) = 0$$

leads to a contradiction, and our theorem is established.

THEOREM III. *If the functions $z_1(x, y)$, $z_2(x, y)$ are defined and continuous on a square Q, and $A_Q(z_1)$ is finite and equal to $\int\int_Q w(z_1)\,dx\,dy$ and*

$$\sqrt{(p_1-p_2)^2+(q_1-q_2)^2} = 0$$

almost everywhere, then either $z_1(x, y)-z_2(x, y)$ is constant or $A_Q(z_2)>A_Q(z_1)$.

Since $p_1 = p_2$ and $q_1 = q_2$ on almost all of Q, we have

$$A_Q(z_1) = \int\int_Q w(z_1)\,dx\,dy = \int\int_Q w(z_2)\,dx\,dy.$$

It is known[9] that

$$A_Q(z_2) \geqq \int\int_Q w(z_2)\,dx\,dy.$$

If the sign $>$ holds, then $A_Q(z_2)>A_Q(z_1)$. If the two quantities are equal, then $z_2(x, y)$ is absolutely continuous, and as in Theorem II we find z_1-z_2 constant.

3. Sequences of surfaces approaching a given surface. We now sharpen somewhat the results of Theorem II.

THEOREM IV. *If the functions $z_1(x, y)$, $z_2(x, y)$ are defined and continuous on a square Q, and there exist positive numbers G, ε, δ, such that*

1) $A_Q(z_1) \leqq G$, $A_Q(z_2) \leqq G$;

2) $\sqrt{(p_1-p_2)^2+(q_1-q_2)^2} \geqq \varepsilon$ *on a set of measure at least equal to* δ;

then

$$A_Q([z_1+z_2]/2) \leqq \tfrac{1}{2}(A_Q(z_1)+A_Q(z_2))-\mu(G, \varepsilon, \delta),$$

where μ is a positive number whose value depends only on G, ε, and δ.

[8] S. Saks, *Sur l'aire des surfaces $z = f(x, y)$*, Acta Szeged, vol. 3 (1927), p. 170.

[9] L. Tonelli, loc. cit. [4], p. 445. S. Saks, loc. cit. [8], p. 175.

Let us again designate by ζ the function $(z_1+z_2)/2$. Since

$$\int\int_Q w(z_1; x, y)\,dx\,dy \leqq A_Q(z_1) \leqq G,$$

we have

3) $w(z_1; x, y) \leqq 4G\delta^{-1}$ on a set of measure $\geqq m(Q) - \delta/4$.

Likewise

4) $w(z_2; x, y) \leqq 4G\delta^{-1}$ on a set of measure $\geqq m(Q) - \delta/4$.

Therefore there exists a set E of measure $\geqq \delta/2$ on which inequalities 2), 3) and 4) are all valid. Consider the function of four variables p_1, q_1, p_2, q_2, defined by

$$\tfrac{1}{2}(\sqrt{1+p_1^2+q_1^2}+\sqrt{1+p_2^2+q_2^2}) - \sqrt{1+([p_1+p_2]/2)^2+([q_1+q_2]/2)^2}.$$

On the bounded closed set of points (p_1, q_1, p_2, q_2) for which

$$\sqrt{1+p_1^2+q_1^2} \leqq 4G\delta^{-1}, \qquad \sqrt{1+p_2^2+q_2^2} \leqq 4G\delta^{-1},$$
$$\sqrt{(p_1-p_2)^2+(q_1-q_2)^2} \geqq \varepsilon,$$

this function is continuous and positive, hence is $\geqq k>0$, where k depends only on G, δ and ε. Moreover, it is never negative, whatever be the values of p_1, q_1, p_2, q_2.

The function of rectangles $\frac{1}{2}(A_R(z_1)+A_R(z_2)) - A_R(\zeta)$ is non-negative, and hence is not less than the integral of its derivative, so that

$$\tfrac{1}{2}(A_Q(z_1)+A_Q(z_2)) - A_Q(\zeta) \geqq \int\int_Q [\tfrac{1}{2}(w(z_1)+w(z_2)) - w(\zeta)]\,dx\,dy$$
$$\geqq \int\int_E [\tfrac{1}{2}(w(z_1)+w(z_2)) - w(\zeta)]\,d(x, y) \geqq k\varepsilon;$$

and since $k\varepsilon$ depends only on G, δ, and ε, our theorem is proved.

THEOREM V. *Let $z(x, y)$ be defined and continuous on a square Q, and $A_Q(z)$ finite; and let $\{z_j(x, y)\}$ be a sequence of functions continuous on Q, such that $\lim z_j = z$ uniformly on Q. Then in order that $\lim A_Q(z_j) = A_Q(z)$ it is necessary (but not sufficient) that $\sqrt{(p_j-p)^2+(q_j-q)^2}$ converge in measure to zero.*

Suppose the theorem false; we can then find a sequence $\{z_j\}$ and a pair of positive numbers ε, δ such that $\lim z_j = z$ uniformly on Q, $\lim A_Q(z_j) = A_Q(z)$, while

$$\sqrt{(p_j-p)^2+(q_j-q)^2} \geqq \varepsilon$$

on a set of measure $\geqq \delta$. There is no loss of generality in assuming that $A_Q(z_j) \leqq A_Q(z)+1$ for every j. Define

$$\zeta_j(x, y) \equiv \tfrac{1}{2}(z(x, y)+z_j(x, y)).$$

By Theorem III we have

$$A_Q(\zeta_j) \leqq \tfrac{1}{2}(A_Q(z) + A_Q(z_j)) - \mu(A_Q(z)+1, \varepsilon, \delta);$$

and since

$$\lim A_Q(z_j) = A_Q(z),$$

this implies

$$\limsup A_Q(\zeta_j) \leqq A_Q(z) - \mu(A_Q(z)+1, \varepsilon, \delta).$$

But by the lower semi-continuity of area, we must have

$$\liminf A_Q(\zeta_j) \geqq A_Q(z);$$

this is the desired contradiction.

It is possible to give a sufficient condition that a sequence $\{z_j(x, y)\}$ of absolutely continuous functions tending uniformly to $z(x, y)$ shall have the property $\lim A_Q(z_j) = A_Q(z)$, but this condition is in a form which appears less applicable than that of Theorem IV. Let us recall that for all surfaces

$$A_Q(z) - \int\int_Q w(z)\, dx\, dy \geqq 0,$$

and vanishes if and only if $z(x, y)$ is absolutely continuous.

THEOREM VI. *Let $z(x, y)$ be defined and continuous on a square Q, and let $\{z_j(x, y)\}$ be a sequence of absolutely continuous functions defined on Q, such that $\lim z_j(x, y) = z(x, y)$ uniformly on Q. Then*

$$\liminf \int\int_Q \sqrt{(p_j-p)^2+(q_j-q)^2}\, dx\, dy \geqq A_Q(z) - \int\int_Q w(z)\, dx\, dy;$$

and

$$\lim A_Q(z_j) = A_Q(z)$$

if

$$\lim \int\int_Q \sqrt{(p_j-p)^2+(q_j-q)^2}\, dx\, dy = A_Q(z) - \int\int_Q w(z)\, dx\, dy.$$

First,

$$\sqrt{(p_j-p)^2+(q_j-q)^2} \geqq w(z_j) - w(z).$$

Integrating and passing to the limit we find that

$$\liminf \int\int_Q \sqrt{(p_j-p)^2+(q_j-q)^2}\, dx\, dy$$

$$\geqq \liminf A_Q(z_j) - \int\int_Q w(z)\, dx\, dy \geqq A_Q(z) - \int\int_Q w(z)\, dx\, dy.$$

Second, assume

$$\lim \int\int_Q \sqrt{(p_j-p)^2+(q_j-q)^2}\, dx\, dy = A_Q(z) - \int\int_Q w(z)\, dx\, dy.$$

Then, since

$$w(z_j) \leqq w(z) + \sqrt{(p_j-p)^2+(q_j-q)^2},$$

we have

$$A_Q(z) \leqq \liminf A_Q(z_j) \leqq \limsup A_Q(z_j)$$

$$\leqq \iint_Q w(z)\, dx\, dy + \lim \iint_Q \sqrt{(p_j-p)^2+(q_j-q)^2}\, dx\, dy = A_Q(z),$$

so that

$$\lim A_Q(z_j) = A_Q(z).$$

4. **Surfaces of least area in the form $z = z(x, y)$.** Suppose that Γ^*: $x = x(t)$, $y = y(t)$ is a simple closed convex curve in the xy plane, and that the curve Γ: $x = x(t)$, $y = y(t)$, $z = z(t)$, projects orthogonally in a one-to-one way on the curve Γ^*. Let D be the point set consisting of Γ^* and the interior of Γ^*. If now we know that a surface $z = z(x, y)$ has minimum area amongst all surfaces bounded by Γ, can we say that it is a minimal surface in the sense of differential geometry? The demonstration is easy if we assume that $z(x, y)$ has continuous first and second partial derivatives, for then it must satisfy the Euler-Lagrange equation for the problem of minimizing $\iint \sqrt{1+p^2+q^2}\, dx\, dy$, and this is the differential equation of minimal surfaces. If we assume merely that p and q are continuous this method fails; the proof for this case was given by Radó[10], and extended by Haar[11] to the case p, q bounded. If we make no hypothesis on p and q, this method in turn fails, or at least cannot be directly extended; but the result remains true, as we shall see in Theorem VIII. We shall however restrict the region D to be a square; the case of general convex regions D is included in Theorem X.

The demonstration consists of two parts, a uniqueness theorem and an existence theorem. The uniqueness theorem is:

THEOREM VII. *Let Q be a square in the xy plane, and Γ a continuous curve which projects orthogonally in a one-to-one way upon the boundary of Q.* [*Let the lower bound A of the areas of surfaces $z = z(x, y)$ bounded by Γ be finite.*][12] *If there exists a continuous surface $z = \zeta(x, y)$ bounded by Γ and such that*

$$A_Q(\zeta) = \iint_Q w(\zeta)\, dx\, dy = A,$$

then $z = \zeta(x, y)$ is the only continuous surface bounded by Γ whose area is A.

[10] T. Radó, *Über den analytischen Charakter der Minimalflächen*, Math. Zeitschrift, vol. 24 (1926), p. 321.

[11] A. Haar, *Über das Plateausche Problem*, Math. Annalen, vol. 97 (1927), pp. 124–158.

[12] We shall see (p. 136, last paragraph) that this hypothesis is always fulfilled.

Let us suppose that $z = z(x, y)$ is a continuous surface bounded by Γ for which $A_Q(z) = A$; we must show that $z(x, y) \equiv \zeta(x, y)$.

A very short proof of this can be given on the basis of previously proved theorems. For by Theorem II we must have

$$\sqrt{(p-\pi)^2+(q-\varkappa)^2} = 0$$

almost everywhere, since otherwise we would have

$$A_Q([z+\zeta]/2) < \tfrac{1}{2}(A_Q(z) + A_Q(\zeta)) = A,$$

which is impossible. Since $A_Q(z) = A_Q(\zeta)$, we find by Theorem III that $z-\zeta$ is constant; and since on the boundary of Q $z-\zeta = 0$, it follows that $z(x, y) \equiv \zeta(x, y)$.

This result can also be established by more elementary methods. Let us construct a sequence of polyhedra $\{z_j(x, y)\}$ $(j = 1, 2, \cdots)$ tending uniformly to $z(x, y)$, such that $\lim A_Q(z_j) = A_Q(z)$. It should be observed that the existence of such a sequence does not follow immediately from the Lebesgue definition of area. For the Lebesgue area is the lower bound, for all possible sequences $\{S_j\}$ of polyhedra

$$S_j: \quad x = x_j(u, v), \qquad y = y_j(u, v), \qquad z = z_j(u, v)$$

tending in the sense of Fréchet to $z = z(x, y)$, of the quantity lim inf (area S_j). We here consider polyhedra which do not possess a simply covered orthogonal projection on the xy plane, and hence cannot immediately conclude that amongst the smaller class of polyhedra representable in the form $z = f(x, y)$ we can find a sequence $\{z_j(x, y)\}$ with the desired properties. That such a sequence can be found follows from a demonstration by Radó.[13] Having chosen such a sequence we define for each $\varepsilon > 0$ the set $E_j(\varepsilon)$ as the set of all points (x, y) for which

$$\sqrt{(p-\pi)^2+(q-\varkappa)^2} \geq \varepsilon.$$

We first show that for each ε $\lim m(E_j(\varepsilon)) = 0$. Suppose the contrary; we can then find an ε such that $\limsup m(E_j(\varepsilon)) > 0$, and therefore we can find a positive δ and a subsequence of the $z_j(x, y)$ (for which we retain the same notation) for which $m(E_j(\varepsilon)) \geq 4\delta$. We can assume without loss of generality that for every j

$$A_Q(z_j) = \iint_Q w(z_j)\, dx\, dy < A+1,$$

so that the point set E_j^* on which $w(z_j) \leq (A+1)\delta^{-1}$ must have measure at least equal to $m(Q)-\delta$. Likewise the set E^* on which $w(\zeta) \leq (A+1)\delta^{-1}$

[13] T. Radó, loc. cit. [4], p. 160.

has measure at least equal to $m(Q) - \delta$. Therefore the sets $E_j(\varepsilon)$, E_j^*, E^* have in common a subset M_j of measure at least 2δ.

If we regard the expression

$$\varphi(p_j, q_j, \pi, \varkappa)$$
$$\equiv \tfrac{1}{2}(\sqrt{1+p_j^2+q_j^2}+\sqrt{1+\pi^2+\varkappa^2})-\sqrt{1+([p_j+\pi]/2)^2+([q_j+\varkappa]/2)^2}$$

as a function of the four variables p_j, q_j, π, $\varkappa$, we see that it is positive and continuous on the bounded closed set defined by the conditions

$$\sqrt{1+p_j^2+q_j^2} \leqq (A+1)\,\delta^{-1}, \qquad \sqrt{1+\pi^2+\varkappa^2} \leqq (A+1)\,\delta^{-1},$$
$$\sqrt{(p_j-\pi)^2+(q_j-\varkappa)^2} \geqq \varepsilon,$$

and on that set must therefore be $\geqq m>0$, where m depends only on A, δ and ε. For all points (x, y) in M_j the derivatives p_j, q_j, π, $\varkappa$ satisfy the conditions; hence

$$\tfrac{1}{2}(A_Q(z_j)+A_Q(\zeta)) - A_Q([z_j+\zeta]/2)$$
$$= \iint_Q \{\tfrac{1}{2}(w(z_j)+w(\zeta)) - w([z_j+\zeta]/2)\}\,dx\,dy$$
$$\geqq \iint_{M_j} \{\tfrac{1}{2}(\sqrt{1+p_j^2+q_j^2}+\sqrt{1+\pi^2+\varkappa^2})$$
$$-\sqrt{1+([p_j+\pi]/2)^2+([q_j+\varkappa]/2)^2}\}\,dx\,dy \geqq \delta m.$$

Hence

$$\limsup A_Q([z_j+\zeta]/2) \leqq \lim \tfrac{1}{2}(A_Q(z_j)+A_Q(\zeta)) - \delta m = A - \delta m.$$

But the surfaces $[z_j+\zeta]/2$ tend uniformly to $[z+\zeta]/2$, and by the lower semi-continuity of area we have

$$A_Q([z+\zeta]/2) \leqq \liminf A_Q([z_j+\zeta]/2) \leqq A - \delta m,$$

which is impossible. We have thus established that for every ε

$$\lim m(E_j(\varepsilon)) = 0.$$

We now show that the assumption $z(x, y) \not\equiv \zeta(x, y)$ leads to a contradiction. Suppose, for concreteness, that $z(x_0, y_0) > \zeta(x_0, y_0)$ at the point (x_0, y_0) interior to Q. (We consider Q as the point set $a_1 \leqq x \leqq b_1$, $a_2 \leqq y \leqq b_2$.) Set

$$z(x_0, y_0) - \zeta(x_0, y_0) = 4d > 0;$$

then there exists a positive h such that

$$z(x_0, y) - \zeta(x_0, y) \geqq 3d$$

for all y such that

$$|y - y_0| \leqq h.$$

For all sufficiently large integers j it is true that

$$|z_j(x, y) - z(x, y)| \leqq d \qquad ((x, y) \text{ on } Q).$$

Hence for all y such that $|y - y_0| \leqq h$ we have

$$z_j(a_1, y) - \zeta(a_1, y) = z_j(a_1, y) - z(a_1, y) \leqq d,$$

and

$$z_j(x_0, y) - \zeta(x_0, y) \geqq z(x_0, y) - \zeta(x_0, y) - |z(x_0, y) - z_j(x_0, y)| \geqq 2d;$$

so that

$$\int_{a_1}^{x_0} (p_j(x, y) - \pi(x, y))\, dx \geqq d,$$

and

$$\int_{y_0-h}^{y_0+h} \int_{a_1}^{x_0} (p_j - \pi)\, dx\, dy \geqq 2hd.$$

A fortiori,

$$\iint_Q \sqrt{(p_j - \pi)^2 + (q_j - \varkappa)^2}\, dx\, dy \geqq 2hd.$$

Let γ represent the number $2hd/5$, and in the definition of $E_j(\varepsilon)$ choose $\varepsilon = \gamma/m(Q)$.

The triangle whose vertices are $(0, 0, 0)$, $(p_j, q_j, 1)$, $(\pi, \varkappa, 1)$ has for the lengths of its sides $w(\zeta)$, $w(z_j)$ and $\sqrt{(p_j - \pi)^2 + (q_j - \varkappa)^2}$; hence we have everywhere

$$w(z_j) - w(\zeta) \geqq \sqrt{(p_j - \pi)^2 + (q_j - \varkappa)^2} - 2w(\zeta).$$

On $Q - E_j(\varepsilon)$,

$$|w(z_j) - w(\zeta)| \leqq \sqrt{(p_j - \pi)^2 + (q_j - \varkappa)^2} < \varepsilon = \gamma/m(Q).$$

We can therefore write

$$\begin{aligned}
&\iint_Q (w(z_j) - w(\zeta))\, dx\, dy \\
&\quad \geqq \iint_{Q-E_j} (w(z_j) - w(\zeta))\, dx\, dy \\
&\qquad + \iint_{E_j} \sqrt{(p_j - \pi)^2 + (q_j - \varkappa)^2}\, dx\, dy - 2\iint_{E_j} w(\zeta)\, dx\, dy \\
&\quad = \iint_{Q-E_j} (w(z_j) - w(\zeta))\, dx\, dy + \iint_Q \sqrt{(p_j - \pi)^2 + (q_j - \varkappa)^2}\, dx\, dy \\
&\qquad - \iint_{Q-E_j} \sqrt{(p_j - \pi)^2 + (q_j - \varkappa)^2}\, dx\, dy - 2\iint_{E_j} w(\zeta)\, dx\, dy \\
&\quad \geqq -\gamma + 5\gamma - \gamma - 2\iint_{E_j} w(\zeta)\, dx\, dy.
\end{aligned}$$

Since $m(E_j(\epsilon))$ tends to zero, for all large enough values of j

$$\int\int_{E_j} w(\zeta)\, dx\, dy \leqq \gamma,$$

and so

$$\int\int_{Q} (w(z_j) - w(\zeta))\, dx\, dy \geqq \gamma.$$

But since

$$\lim \int\int_{Q} w(z_j)\, dx\, dy = A$$

and

$$\int\int_{Q} w(\zeta)\, dx\, dy = A$$

this is absurd. Hence the assumption $z(x, y) \not\equiv \zeta(x, y)$ leads to a contradiction, and so we must have $z(x, y) \equiv \zeta(x, y)$, proving the theorem.

If now it be possible to assert that there exists a surface $z = \zeta(x, y)$ which is minimal in the sense of differential geometry, is bounded by Γ and has area $A_Q(\zeta) = A$, we could then state immediately that every continuous surface $z = z(x, y)$ bounded by Γ and having area A is analytic; for then by Theorem VII we have $z(x, y) \equiv \zeta(x, y)$, and $\zeta(x, y)$ is an analytic function[14] of x and y. We can in fact make such an assertion, thanks to the researches of Radó, who established the following important result:

For any Jordan curve Γ in the $x\ y\ z$ space, the problem of the least area has infinitely many solutions. Among these solutions, there is always an analytic surface, and if Γ bounds at least one continuous surface with a finite area, then there is even a minimal surface solving the problem of the least area.[15]

The surfaces here considered are in parametric form. For our case we easily exhibit a surface bounded by Γ and of finite area, hence the least area A^* of all surfaces of the type of the circle bounded by Γ is finite. Hence there exists a minimal surface $x = f_1(u, v)$, $y = f_2(u, v)$, $z = f_3(u, v)$ bounded by Γ and with area A^*. We have now only to apply the theorem[16] that whenever Γ has a simply covered orthogonal projection on the $x\ y$ plane, the minimal surface bounded by Γ can be represented in the form $z = \zeta(x, y)$. Since $z = \zeta(x, y)$ has least area amongst all continuous surfaces of the type of the circle bounded by Γ, it has area A.

[14] T. Radó, loc. cit. [10].

[15] T. Radó, *The problem of the least area and the problem of Plateau*, Math. Zeitschrift, vol. 32, p. 763.

[16] T. Radó, *Some remarks on the problem of Plateau*, Proceedings of the National Academy of Sciences, vol. 16 (1930), p. 247.

(This shows incidentally that the hypothesis that A is finite, used in Theorem VII, is always satisfied.)

Combining these results, we have

THEOREM VIII. *Let Q be a square in the $x\,y$ plane, and Γ a continuous curve which projects orthogonally in a one-to-one way upon the boundary of Q. Then if the surface $z = z(x, y)$ is bounded by Γ and has the least possible area amongst all surfaces $z = f(x, y)$ bounded by Γ, it is analytic, and moreover it is a minimal surface in the sense of differential geometry.*

It would be interesting to have an independent proof of this theorem which would not require both the uniqueness theorem and the existence theorem used in the preceding pages.

Anticipating the extension of Theorem VIII to general convex regions, we can sum up the results thus far in

THEOREM IX. *Let the closed curve Γ have for its orthogonal projection on the $x\,y$ plane a simply covered convex curve Γ^*, and let D be the point set consisting of Γ^* and its interior. Then*

1) *there exists a unique*[17] *minimal surface bounded by Γ, and this surface can be represented in the form $z = \zeta(x, y)$ ((x, y) on D);*
2) *amongst all continuous surfaces $z = f(x, y)$ ((x, y) on D) bounded by Γ there exists a unique surface $z = z(x, y)$ having least area;*
3) *these two surfaces coincide.*

5. Surfaces of least area in parametric form. Let us now turn our attention to surfaces defined by three continuous functions $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$, $u^2+v^2 \leqq 1$. Suppose that we are given a closed curve Γ, such that the lower bound of the areas of all such surfaces bounded by Γ is A. It is then definitely *not* true that every surface $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$, $u^2+v^2 \leq 1$, bounded by Γ and having area A, is a minimal surface, as has been pointed out by Lebesgue;[18] for given any one such surface we can alter it by adding a spine without changing its area, and such a surface cannot be minimal. Moreover, examples of such surfaces with spines can be given possessing any desired number of continuous partial derivatives, so that addition of hypotheses as to the number of derivatives of x, y, z is of no avail. It is, however, sufficient[19] to assume that the first partial derivatives are continuous, and that $EG-F^2 \neq 0$. The effect of this last hypothesis is to make a sufficiently small portion of the surface a topological image of a portion of the uv plane, and in fact to assure a one-to-one orthogonal projection on a certain

[17] T. Radó, loc. cit. [16], p. 247.

[18] H. Lebesgue, *Intégrale, Longueur, Aire*, Annali di Matematica, Serie III[a], vol. 7 (1902), pp. 231–359.

[19] T. Radó, loc. cit. [10].

plane; that is, it introduces a condition of topological regularity. We will now show that this topological condition is all that is needed, without reference to derivatives. It seems probable that every portion of such a surface which is a topological image of a region of the uv plane is minimal; this I have not been able to prove. However, it is possible to prove

THEOREM X. *Let the surface* S: $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$, $u^2 + v^2 \leqq 1$, *be bounded by the curve* Γ, *and have a finite area which is the least area amongst all surfaces bounded by* Γ. *Then if a portion of* S *have a simply covered orthogonal projection on some plane, that portion of* S *is a minimal surface in the sense of differential geometry.*

There is clearly no loss of generality in assuming the plane to be the xy plane. Our hypothesis can then be restated in the form that a region of the uv plane is carried by the equations $x = x(u, v)$, $y = y(u, v)$ into a region of the xy plane. Let Q^* be a square $a_1^* \leqq x \leqq b_1^*$, $a_2^* \leqq y \leqq b_2^*$ lying in this region; to each point (x, y) in Q^* there corresponds one and only one (u, v), and to Q^* corresponds a one-to-one continuous image P^*. Then the surface S^*: $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$, (u, v) *in* P^*, can be represented in the form $z = f(x, y)$ ((x, y) *in* Q^*). Since S^* is a portion of S, its area is finite; hence for almost all x in the interval (a_1^*, b_1^*) the function $f(x, y)$ is of limited total variation in y, and for almost all y in the interval (a_2^*, b_2^*) the function $f(x, y)$ is of limited total variation in x. We can therefore find a square Q: $(a_1 \leqq x \leqq b_1, a_2 \leqq y \leqq b_2)$ interior to Q^*, such that the curve Γ_0: $z = f(x, y)$ ((x, y) *on boundary of* Q) is rectifiable. Utilizing certain results given by Lebesgue[20] we find that the area of the surface $z = f(x, y)$ ((x, y) *on* Q) is least amongst all surfaces $z = \varphi(x, y)$ ((x, y) *on* Q) bounded by Γ_0. (The point of importance here is to show that if a surface be divided in two portions by a rectifiable curve, the area of the whole surface is the sum of the areas of the two portions. This result, which seems intuitively evident, can be deduced from Lebesgue's thesis; it would seem to be of some interest to give an independent demonstration of that portion which is needed here.) Hence, by Theorem VIII, $z = f(x, y)$ is minimal and consequently analytic, and the theorem is proved.

OHIO STATE UNIVERSITY.
February 28, 1931.

[20] H. Lebesgue, loc. cit.

THE EXPANSION OF ANALYTIC FUNCTIONS IN GENERALIZED LAMBERT SERIES.[1]

By J. M. Feld.

1. **Introduction.** The purpose of this paper is to show that any analytic function $f(z) = \sum_0^\infty c_n z^n$ can be represented in the form

$$f(z) = b_0 + \sum_1^\infty b_n \frac{a_n z^n}{1 - a_n z^n} \tag{1}$$

in some region about the origin, where the b_n $(n \neq 0)$ are a set of preassigned constants with the restrictions that for all values of n except zero, $|b_{n+1}| \geqq |b_n| \geqq 1$ and that $|b_n|^{1/n} < M$ where M is a positive constant. From this it will follow that any function analytic and equal to unity in the neighborhood of the origin can be expanded in the form of an infinite product

$$\prod_1^\infty (1 - a_n z^n)^{-b_n/n}.$$

The product expansion is a generalization of an expansion found by J. F. Ritt.[2]

2. Let $f(z)$ be formally represented in the form (1). Expanding the right member of (1) in a power series and equating coefficients we get the recursion formulas

$$\begin{aligned} c_0 &= b_0, \\ c_1 &= b_1 a_1, \\ c_2 &= b_1 a_1^2 + b_2 a_2, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ c_n &= b_1 a_{d_0}^n + b_{d_1} a_{d_1}^{n/d_1} + \cdots + b_{d_t} a_{d_t}^{n/d_t} + b_n a_n, \end{aligned} \tag{2}$$

where $d_0 = 1, d_1, \cdots, d_t, n$ are the divisors of n and d_t is the greatest divisor of n smaller than n.

From (2) we get

$$|a_n| \leqq \frac{1}{|b_n|}\left[|c_n| + \sum_{i=0}^{t} |b_{d_i} a_{d_i}^{n/d_i}|\right]. \tag{3}$$

---

[1] Received May 12, 1930, September 5, 1930, and June 24, 1931.

[2] Mathematische Zeitschrift, 32 (1930), pp. 1–3.

Let $|b_{n_1} a_{n_1}^{n/n_1}|$ be the greatest of the terms appearing after the $\sum$ in (3). Then

$$|a_n| \leqq \frac{1}{|b_n|}\left[|c_n| + \frac{n}{2}|b_{n_1} a_{n_1}^{n/n_1}|\right]$$

since $t+1$ cannot exceed $n/2$.

If

$$|c_n| \geqq \frac{n}{2}|b_{n_1} a_{n_1}^{n/n_1}|, \qquad |a_n| \leqq \frac{1}{|b_n|}[2|c_n|] \leqq 2|c_n|,$$

and

$$|a_n|^{1/n} \leqq 2^{1/n}|c_n|^{1/n} < 2r$$

where r is the least upper bound of the set $|c_1|, |c_2|^{1/2}, \cdots, |c_n|^{1/n}, \cdots$.

If

$$|c_n| < \frac{n}{2}|b_{n_1} a_{n_1}^{n/n_1}|, \qquad |a_n| < \frac{1}{|b_n|} \cdot n\,|b_{n_1} a_{n_1}^{n/n_1}|,$$

$$|a_n|^{1/n} < \left|\frac{b_{n_1}}{b_n}\right|^{1/n} n^{1/n}|a_{n_1}|^{1/n_1} \leqq n^{1/n}|a_{n_1}|^{1/n_1}.$$

By the same reasoning we have either

$$|a_{n_1}|^{1/n_1} \leqq 2r$$

or

$$|a_{n_1}|^{1/n_1} < n_1^{1/n_1}|a_{n_2}|^{1/n_2}$$

where n_2 is a divisor of n_1 less than n_1. Repeating this procedure we finally arrive at

$$|a_n|^{1/n} < n^{1/n} n_1^{1/n_1} \cdots n_k^{1/n_k}|a_{n_{k+1}}|^{1/n_{k+1}} \tag{4}$$

where $n_{k+1} \geqq 1$ and $n_k \geqq 2$. From the recursion formulas (2) we find $|a_1| < 2r$. We can therefore always obtain an inequality (4) in which the last factor

$$|a_{n_{k+1}}|^{1/n_{k+1}} < 2r.$$

From (4) we obtain

$$\frac{1}{n}\log|a_n| < \frac{1}{n}\log n + \frac{1}{n_1}\log n_1 + \cdots + \frac{1}{n_k}\log n_k + \log(2r).$$

Since $n_p \geqq 2n_{p+1}$ we have for $n_k \neq 3$

$$\frac{1}{n}\log|a_n| < \log(2r) + \sum_{k=1}^{k=\infty}\frac{1}{2^k}\log 2^k = \log(2r) + \log 4 = \log 8r.$$

If $n^k = 3$ we get

$$\frac{1}{n}\log|a_n| < \log(2r) + \sum_{k=0}^{k=\infty} \frac{1}{3\cdot 2^k}\log 3\cdot 2^k$$
$$= \log(2r) + \frac{2}{3}\log 3 + \frac{1}{3}\log 4 < \log 8r.$$

Thus we always have $|a_n|^{1/n} < 8r = R$.

For $|z| < \frac{1}{R}$

$$\sum_{n=1}^{n=\infty} \left| b_n \frac{a_n z^n}{1-a_n z^n} \right| < \sum |b_n| \cdot \frac{R^n |z|^n}{1-|a_n z^n|} < \sum |b_n| \frac{|Rz|^n}{1-|Rz|^n}.$$

But $\sum |b_n| \frac{|Rz|^n}{1-|Rz|^n}$ is a Lambert series and converges in the same region in which $\sum_1^\infty |b_n| w^n$ (where $w = |Rz|$) converges. If $\sum |b_n| w^n$ converges uniformly for $w < l$, (1) will converge uniformly for $|z|$ less than the smaller of the numbers $1/R$ and l/R.

Writing (1) in the form

$$f(z) - b_0 = -z \sum_1^\infty \left(-\frac{b_n}{n} \cdot \frac{a_n n z^{n-1}}{1-a_n z^n} \right)$$

we get

$$F(z) = -\int_0^z \frac{f(z)-b_0}{z}\,dz = \sum_1^\infty \left[-\frac{b_n}{n}\log(1-a_n z^n) \right]$$
$$= \log \prod_1^\infty (1-a_n z^n)^{-b_n/n}$$

and

$$\Phi(z) = e^{F(z)} = \prod_1^\infty (1-a_n z^n)^{-b_n/n}$$

where $\Phi(z)$ is regular in the neighborhood of the origin. This reduces to the product expansion obtained by J. F. Ritt if $b_n = -n$.

We have therefore proved

THEOREM 1. *Given an infinite sequence of constants b_n, $n > 0$ such that $|b_{n+1}| \geqq |b_n| \geqq 1$, $|b_n|^{1/n} < M$, any function $f(z)$, analytic in the neighborhood of the origin, can be uniquely represented in the neighborhood of the origin as a uniformly convergent series of the form*

$$b_0 + \sum_{n=1}^{n=\infty} b_n \frac{a_n z^n}{1-a_n z^n}.$$

THEOREM 2. *Given an infinite sequence of constants* b_n, $n>0$ *such that* $|b_{n+1}| \geq |b_n| \geq 1$, $|b_n|^{1/n} < M$, *then any function* $\Phi(z)$ *analytic in a neighborhood of the origin and equal to unity at the origin can be represented as an absolutely convergent infinite product* $\prod_1^\infty (1 - a_n z^n)^{-b_n/n}$ *in some neighborhood of the origin.*

If the a_n are preassigned a unique set of b_n can be found for which (1) will converge uniformly in some region about the origin. This follows from theorems due to Pincherle,[3] Izumi,[4] Okada[5] and Narumi.[6]

[3] Pincherle, Memorie della Accademia delle Scienze dell' Instituto di Bologna, (4) 3 (1882) 151–180.

[4] Izumi, Tôhoku Math. Journal, 28 (1927) 97.

[5] Okada, id. 27 (1922–3) 324.

[6] Narumi, id. 30 (1928–9) 441.

NEW YORK, N. Y.

ARITHMETICAL PERIODICITIES OF BESSEL FUNCTIONS.[1]

By D. H. Lehmer.[2]

Introduction. The arithmetical theory of the solutions of a difference equation depends on the order of the equation and on the nature of its coefficients. It is possible to associate such theories with the functions of analysis which satisfy the type of difference equation under discussion. Thus we associate the arithmetical theories of recurring series[3] with exponential, circular, elliptic, or hyper-elliptic functions according as the order of recurrence is 1, 2, 3, or more.[4] Again the topics allied to Wilson's theorem give arithmetical properties of the Gamma function. From this point of view the present paper is that part of the arithmetic of Bessel functions which deals with their periodicity with respect to a modulus. The difference equation under consideration is

$$U_{n+1} = (an+b)\,U_n + c\,U_{n-1} \tag{1}$$

where a, b, and c are integers.

The first part of the paper (§§ 1, 2, 3) shows the connection between U_n and Bessel functions. The latter part deals with U_n modulo M, the main theorem being Theorem 6. A part of this theory ($c = 1$) has been considered by D. N. Lehmer[5] in discussing the divisibility of the numerators and denominators of a certain set of convergents of special Hurwitzian continued fractions. By contracting[6] the continued fraction, these numbers may be shown to satisfy (1).

Comparing what follows with the theory of recurring series, one sees that the present paper is distinctly *non-algebraic*. In fact there is no "characteristic equation" and consequently no associated algebraic number field in which to study the solutions of (1). Moreover, neither the formulas giving the complete solution of the difference equation nor the actual prime factors of the modulus are essential to the discussion.

1. **The complete solution of (1).** It is often convenient to introduce the number

$$\alpha = b/a$$

[1] Received February, 1931.

[2] National Research Fellow.

[3] That is, linear difference equations with constant coefficients.

[4] This has been done by Lucas: Amer. Jour. Math., 1 (1878), pp. 184–239, pp. 289–321.

[5] Amer. Jour. Math., 40 (1918), 375–390, Proc. Nat. Acad. Sciences, 4 (1918), pp. 214–218.

[6] Perron: Die Lehre von den Kettenbrüchen (Berlin, 1913), pp. 134, 202.

and write (1) in the form

$$U_{n+1} = a(n+\alpha)\,U_n + c\,U_{n-1}.$$

The function U_n is completely determined when a, α, c, U_0, U_1 are given. That is

$$U_n = U_n(a, \alpha, c; U_0, U_1).$$

The two functions

$$U_n(a, \alpha, c; 0, 1) \quad \text{and} \quad U_n(a, \alpha, c; 1, 0)$$

are of special interest and are designated by $S_n(a, \alpha, c)$ and $T_n(a, \alpha, c)$ respectively. In terms of these functions the general U_n is given by

$$U_n = U_0\,T_n + U_1\,S_n. \tag{2}$$

As for the actual formulas for S_n and T_n, they are polynomials in c/a^2. In fact it is not difficult to verify that for $n>0$

$$S_n(a, \alpha, c) = a^{n-1} \sum_{k=0}^{[(n-1)/2]} (c/a^2)^k \binom{n-k-1}{k} \frac{\Gamma(n-k+\alpha)}{\Gamma(k+\alpha+1)}$$

and for $n>1$

$$T_n(a, \alpha, c) = a^n \sum_{k=0}^{[(n-2)/2]} (c/a^2)^{k+1} \binom{n-k-2}{k} \frac{\Gamma(n-k+\alpha)}{\Gamma(k+\alpha+2)}.$$

Substituting these expressions in (2) we obtain the complete solution of the difference equation (1).

2. **Solution of (1) by Bessel's functions.** If, in the well known recurrence formula for Bessel's functions

$$J_{m+1}(z) = \frac{2m}{z} J_m(z) - J_{m-1}(z), \tag{3}$$

we replace m by $n+\alpha$ and z by $2\sqrt{-c}/a$ and then multiply both members by $(-c)^{(n+1)/2}$ we obtain

$$(-c)^{(n+1)/2}\, J_{n+\alpha+1}\left(\frac{2\sqrt{-c}}{a}\right)$$

$$= a(n+\alpha)\,(-c)^{n/2}\, J_{n+\alpha}\left(\frac{2\sqrt{-c}}{a}\right) + c(-c)^{(n-1)/2}\, J_{n+\alpha-1}\left(\frac{2\sqrt{-c}}{a}\right).$$

Hence the function J^* defined by

$$J_n^*(a, \alpha, c) = (-c)^{n/2}\, J_{n+\alpha}\left(\frac{2\sqrt{-c}}{a}\right)$$

is a particular solution of (1).

In the same way it is seen that the function

$$Y_n^*(a, \alpha, c) = (-c)^{n/2} Y_{n+\alpha}\left(\frac{2\sqrt{-c}}{a}\right)$$

where $Y_m(z)$ is the second solution of Bessel's equation,[7] satisfies (1). The complete solution in terms of Bessel's functions is a linear combination of Y^* and J^*.

3. **Lommel's polynomials.** The particular solution J^* of (1) may be expressed in terms of S and T:

$$J_n^* = J_0^* T_n + J_1^* S_n$$

or

$$(-c)^{n/2} J_{n+\alpha}\left(\frac{2\sqrt{-c}}{a}\right)$$

(4)

$$= J_\alpha\left(\frac{2\sqrt{-c}}{a}\right) T_n(a, \alpha, c) + \sqrt{-c}\, J_{\alpha+1}\left(\frac{2\sqrt{-c}}{a}\right) S_n(a, \alpha, c).$$

By successive application of formula (3) it is possible to express $J_{m+\nu}$ linearly in terms of J_ν and $J_{\nu+1}$. The coefficients entering into this expression are known as Lommel's polynomials.[8] In fact we have the relation[9]

$$J_{m+\nu}(z) = R_{m-1,\nu+1}(z) J_\nu(z) - R_{m-2,\nu+2}(z) J_{\nu+1}(z).$$

If we multiply both members by $(-c)^{n/2}$ and replace (m, ν, z) by $(n, \alpha, 2\sqrt{-c}/a)$ we obtain the following relations, after comparing with (4).

$$S_n(a, \alpha, c) = (-c)^{(n-1)/2} R_{n-1,\alpha+1}\left(\frac{2\sqrt{-c}}{a}\right),$$

$$T_n(a, \alpha, c) = -(-c)^{n/2} R_{n-2,\alpha+2}\left(\frac{2\sqrt{-c}}{a}\right).$$

4. **The general solution U_n modulo M.** In what follows we consider U_n modulo M, any number prime to a. In all congruences where the modulus is not indicated it is understood to be M.

THEOREM 1. *Let $b/a = \alpha \equiv -\varrho \pmod{M}$. Then if $k \geqq 0$ is an integer*

$$U_{\varrho+k} \equiv c^k U_{\varrho-k} \pmod{M}.$$

Proof. The theorem is trivial for $k = 0$. Supposing it true for $k < h$ we may prove it for $k = h$ as follows. Since $a\varrho + b \equiv 0$ we have from (1)

[7] See for example Watson's, *Bessel's functions*, Cambridge 1922, chap. 3.

[8] Lommel, Math. Annalen 4, pp. 108–116 (1871).

[9] Watson, loc. cit., chap. 9.

$$\begin{aligned} U_{\varrho+h} &= [a(\varrho+h-1)+b]\,U_{\varrho+h-1}+cU_{\varrho+h-2} \\ &\equiv a(h-1)c^{h-1}\,U_{\varrho-h+1}+c^{h-1}\,U_{\varrho-h+2} \\ &\equiv c^{h-1}\{U_{\varrho-h+2}-[a(\varrho-h+1)+b]\,U_{\varrho-h+1}\} \\ &\equiv c^h\,U_{\varrho-h}. \end{aligned}$$

This completes the induction. Since ϱ has an infinity of values differing by multiples of M we may state Theorem 1 in the form

$$U_{rM+\varrho+k} \equiv c^k\,U_{rM+\varrho-k} \tag{5}$$

where r and k are any integers $\geqq 0$. In case $c \equiv 1$, (5) states that the U's taken modulo M are palindromic with centers at ϱ, $M+\varrho$, $2M+\varrho$, $\cdots$.

5. First Periodicity Theorem.

THEOREM 2. *If n is any integer, then*

$$U_{n+2M} \equiv c^M\,U_n \pmod{M}.$$

Proof. Let $n=(r-1)M+\varrho+k$. Then by (5)

$$\begin{aligned} U_{n+2M} &= U_{rM+\varrho+M+k} \equiv c^{M+k}\,U_{(r-1)M+\varrho-k} \\ &\equiv c^M\,U_{(r-1)M+\varrho+k} \equiv c^M\,U_n, \end{aligned}$$

which is the theorem. Applying this theorem r times we have

$$U_{n+2rM} \equiv c^{rM}\,U_n. \tag{6}$$

This gives us the following

THEOREM 3. *If ac is prime to M and if e is the exponent to which c belongs (modulo M), then $U_n \pmod{M}$ is periodic of period $2eM$.*

THEOREM 4. *Let λ and μ be any integers. If the congruence*

$$U_{n+\lambda M} \equiv \mu U_n \pmod{M} \tag{7}$$

is true for any two consecutive values of n then it is true for all values of n.

Proof. Let $n=\nu-1$ and $n=\nu$ be the values of n for which (7) is true. Then it is sufficient to show that (7) is true for $n=\nu+1$. We have in fact

$$\begin{aligned} U_{\nu+1+\lambda M} &= [a(\nu+\lambda M)+b]\,U_{\nu+\lambda M}+cU_{\nu-1+\lambda M} \\ &\equiv \mu(a\nu+b)\,U_\nu+\mu cU_{\nu-1} \equiv \mu U_{\nu+1}. \end{aligned}$$

Hence the theorem.

6. Second Periodicity Theorem.

It is now time to come to an understanding about the terms of the sequence U_n $(n=0, 1, \cdots)$. It is clear that no great generality is lost in supposing that U_0 and U_1 (and hence U_n) are integers. Moreover any factor common to U_0 and U_1 will be common to all U_n. This extraneous factor may be removed without

changing the essential character of the sequence U. We therefore require in what follows that U_0 and U_1 be coprime integers. All possible solutions of (1) are then to be obtained from the set just abstracted by multiplying every U_n by an arbitrary constant which may or may not be an integer. With this fundamental hypothesis we may proceed to strengthen Theorem 3.

THEOREM 5. *If ac is prime to M, then the proper period of $U_n \pmod M$ is divisible by M.*

Proof. Let π be the proper period in question. Then we will have in particular

$$U_{\pi+1} \equiv U_1$$

that is

$$(\pi a + b) U_\pi + c U_{\pi-1} \equiv b U_0 + c U_{-1}$$

or

$$\pi a U_0 \equiv 0 \pmod M. \tag{8}$$

From the condition

$$U_{\pi+2} \equiv U_2$$

we derive in the same way

$$\pi a U_1 \equiv 0. \tag{9}$$

Now since U_0 and U_1 are coprime, integers x and y exist for which

$$U_0 x + U_1 y = 1.$$

Multiplying (8) by x and (9) by y and adding we get

$$\pi a \equiv 0 \pmod M.$$

Since a is prime to M, the theorem follows.

THEOREM 6. *If ac is prime to M the proper period of $U_n \pmod M$ is the product of the L. C. M. of e and M by some divisor of twice their G. C. D.*

Proof. By Theorem 3 there exists an integer k such that

$$k\pi = 2eM \tag{10}$$

π being the proper period of U_n. By Theorem 5, π is divisible by M. Hence k is a divisor of $2e$. If in (6) we set $rM = \pi$ we obtain

$$U_{n+2\pi} \equiv c^\pi U_n \equiv U_n,$$

since 2π is a period. Therefore π is a multiple of e and it follows from (10) that k is a divisor of $2M$. Hence k is a divisor of $2(e, M)$, writing

$$k\delta = 2(e, M)$$

we have from (10)

$$\pi = \frac{eM}{(e, M)}\delta = [e, M]\delta \tag{11}$$

where δ is some divisor of $2(e, M)$. Hence the theorem.

COROLLARY 1. *If e is prime to M, then either $\pi = eM$, or $\pi = 2eM$.*

COROLLARY 2. *If $M = p^\alpha$ is a power of a prime, then $\pi = \eta e_0 p^{\alpha+\nu}$ where $\eta = 1$ or 2 and $0 \leqq \nu \leqq \alpha - 1$ and e_0 is the exponent to which c belongs modulo p.*

Theorem 6 gives the upper and lower bounds of π considered as a function of a, b, U_0, and U_1, with c (and hence e) and M fixed. That is

$$2eM \geqq \pi \geqq [e, M].$$

The example in § 7 shows that these limits are actually attained. To illustrate the case in which both inequalities hold, consider the sequence

$$1, 1, 7, 65, 859, 14733, 311111, \cdots$$

in which

$$U_{n+1} = (4n+1)\,U_n + 2\,U_{n-1}.$$

Let $M = 21$, then $e = 6$ is the exponent to which 2 belongs. According to Theorem 3, π might be any of the 18 divisors of 252. Theorem 5 rules out all but

$$21, 42, 63, 84, 126, 252$$

while Theorem 6 rules out 21 and 63. As a matter of fact $\pi = 84$.

7. **The special case $a \equiv \pm 2b$.** Since the Bessel functions of order $n+\alpha$ are elementary when $\alpha = \pm 1/2$, one might expect that the case $a \equiv \pm 2b$ should also be simple. At any rate such is the case. Since $(M, a) = 1$, M is odd in this section.

THEOREM 7. *Let $a \equiv -2\varepsilon b$, where $\varepsilon = \pm 1$ and where M is prime to a. Then for any integer λ,*

$$U_{M+\lambda} \equiv U_{\varepsilon-\lambda}\, c^{\lambda+(M-\varepsilon)/2} \pmod{M}. \tag{12}$$

Proof. In Theorem 1 set $\varrho = (M+\varepsilon)/2$, $k = \lambda + (M-\varepsilon)/2$, we have then

$$\varrho + k = M + \lambda, \quad \text{and} \quad \varrho - k = \varepsilon - \lambda.$$

Hence the theorem.

Let us next consider under what conditions the proper period π will have the smallest possible value, namely M. It is necessary in the first place to make $e = 1$, because any larger value of e would be even and hence $[e, M] > M$ since M is odd. In other words $c \equiv 1$. It is next necessary that

$$U_M \equiv U_0.$$

In view of (12) this becomes

$$U_\varepsilon \equiv U_0. \tag{13}$$

By Theorem 4 it is finally sufficient to impose the condition

$$U_{M+1} \equiv U_1.$$

That is

$$U_{\varepsilon-1} \equiv U_1. \tag{14}$$

We shall now show that (13) implies (14). In fact this is obvious when $\varepsilon = 1$. When $\varepsilon = -1$, (13) becomes

$$U_1 \equiv (b+1)\,U_0 \tag{15}$$

while

$$U_{\varepsilon-1} = U_{-2} \equiv U_0 - (b-a)\,U_{-1} \equiv (b+1)\,U_0 \equiv U_1.$$

Hence we have

THEOREM 8. *If $a \equiv -2\varepsilon b$, $\varepsilon = \pm 1$, then in order that the proper period π should be as small as possible it is necessary and sufficient that $c \equiv 1$ and that $U_\varepsilon \equiv U_0$.*

Example. The convergents to the continued fraction

$$1 + \frac{2}{1} + \frac{1}{6} + \frac{1}{10} + \frac{1}{14} + \frac{1}{18} + \cdots$$

for the Naperian base are

n	0	1	2	3	4	5	6	7	$\cdots$,
A_n	1	3	19	193	2721	49171	1084483	28245729	
B_n	1	1	7	71	1001	18089	398959	10391023	$\cdots$.

The numerators and denominators both satisfy the recurrence

$$U_{n+1} = (4n+2)\,U_n + U_{n-1}.$$

In this case $a = 4$, $b = 2$, $c = 1$, $\varepsilon = -1$. Moreover $A_1 = 3 = (b+1)A_0$, hence for an arbitrary odd integer M the condition (13) is satisfied so that modulo M the numerators have M for a proper period. The denominators on the other hand have a proper period of $2M$. In fact it follows from Theorem 7 that

$$B_M \equiv B_{-1} \equiv -B_0; \qquad B_{M+1} \equiv B_{-2} \equiv -B_1.$$

Hence by Theorem 4 we conclude that[10]

$$B_{M+n} \equiv -B_n.$$

For example if $M = 7$ we have

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\cdots$,
A_n	1	3	5	4	5	3	1	1	3	5	4	5	3	1	1	3	$\cdots$,
B_n	1	1	0	1	0	1	1	6	6	0	6	0	6	6	1	1	$\cdots$.

[10] These simple facts about the convergents to the Naperian base furnished invaluable checks in computing A_n and B_n. See American Journal of Mathematics, vol. 48 (1926), p. 140.

8. **Divisibility of S_n and T_n.** It is natural to inquire for what values of n is U_n divisible by M. This question does not have a simple answer when U_0 and U_1 are not specified.[11] But when $(U_a, U_1) = (0, 1)$ or $(1,0)$ we have the following theorem which is an immediate consequence of (5) and (6).

THEOREM 9. *If r is any integer and if $a\varrho + b \equiv 0 \pmod M$, then*

$$S_{2rM} \equiv S_{2(rM+\varrho)} \equiv T_{2rM+1} \equiv T_{2(rM+\varrho)+1} \equiv 0 \pmod M.$$

Other theorems similar to Theorem 8 may be written down when a given term U_ν is known to be zero or congruent to zero. On the other hand there may be an n not of the form $2rM$ or $2(rM+\varrho)$ for which $S_n \equiv 0$. Although such an event is practically unpredictable we may infer from it the existance of infinitely many other multiples of M in the sequence U_n. That these unpredictable cases must occur for some values of M is seen from the fact that the terms of S_n increase much more rapidly than n itself. The following table, giving the decompositions into prime factors of $S_n(1, 0, -1)$, $n \leq 14$, shows that there is no simple law of "apparition" or "repetition" of primes.

n	S_n	n	S_n		n	S_n	
0	0	5	7		10	90865	$= 5 \cdot 17 \cdot 1069$
1	1	6	33	$= 3 \cdot 11$	11	898409	$=$ prime
2	1	7	191	$=$ prime	12	9791634	$= 2 \cdot 3 \cdot 1631939$
3	1	8	1304	$= 2^3 \cdot 163$	13	116601199	$= 11 \cdot 13 \cdot 29 \cdot 31 \cdot 907$
4	2	9	10241	$= 7^2 \cdot 11 \cdot 19$	14	1506023953	$= 7 \cdot 215146279.$

In this case $\varrho = 0$ so by Theorem 8 we can assert that the prime p will occur as a factor of S_{2p}. A glance at the above table shows how weak this assertion is. It also indicates that a complete theory of the divisors of S_n is practically out of the question.

[11] In the same way the *general* recurring series $U_{n+1} = PU_n - QU_{n-1}$ has no simple divisibility theorem.

CALIFORNIA INSTITUTE OF TECHNOLOGY.

A NOTE ON HYPERGEOMETRIC FUNCTIONS OF TWO VARIABLES.[1]

By Thomas H. Rawles.

The series

$$F(\alpha, \beta, \gamma, x) = 1 + \frac{\alpha \cdot \beta}{\gamma \cdot 1} x + \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)1 \cdot 2} x^2 + \cdots$$

was called by Gauss[2] the hypergeometric series, and the second order linear differential equation of which it is a solution is known as the hypergeometric equation.

Generalizations of the Gaussian function are due to Pochhammer[3] and to Goursat.[4] The former defines as a hypergeometric function of the kth order one which has, among others, the following properties:

1) It is the general solution of a linear differential equation of the kth order,
2) It has $k+1$ distinct singular points of which one is at infinity.

The two other conditions prescribed are concerned with the development of the function in the neighborhoods of the singular points.[5]

From this definition Pochhammer derives the differential equation satisfied by the function, and then shows how the solution of this equation may be expressed by means of the definite integral

$$y = \int_g^h (u - a_1)^{b_1 - 1} (u - a_2)^{b_2 - 1} \cdots (u - a_k)^{b_k - 1} (u - x)^{\lambda - 1} \, du,$$

where g and h are any two of the quantities $a_1, a_2, \cdots, a_k, x$. This reduces to the Gaussian function if we put $k = 2$, $a_1 = 0$, $a_2 = 1$, $b_1 = \alpha - \gamma + 1$, $b_2 = \gamma - \beta$, $\lambda = 1 - \alpha$.

According to Goursat[6] a hypergeometric function is one defined by the power series $\sum_0^\infty a_n x^n$, where the ratio a_{n+1}/a_n is a rational function of n.

---

[1] Received January 26, 1931.

[2] Gauss, Disquisitiones generales circa seriem infinitam $1 + \frac{\alpha \cdot \beta}{\gamma \cdot 1} x + \cdots$.

[3] Pochhammer, Über hypergeometrische Funktionen $n$ter Ordnung, Journal für die reine und angewandte Mathematik, vol. LXXI, 1869, pp. 316–52.

[4] Goursat, Sur les fonctions hypergéométriques d'ordre supérieur, Annales de l'École Normale, vol. XII, 1883, pp. 261–86, pp. 395–430.

[5] Pochhammer, loc. cit. p. 324.

[6] Goursat, loc. cit. p. 261.

The development of the theory of hypergeometric functions of two and more variables is due principally to Appell.[7] Besides contributing numerous memoirs to their study he has, in collaboration with Kampé de Feriet, devoted an elaborate treatise to the investigation of their properties.

The viewpoint of these authors follows that of Goursat in that they regard as a hypergeometric series one in which the ratio of adjacent terms is a fixed rational function of the rank of one of them. That is to say if $z = \sum_0^\infty \sum_0^\infty a_{mn} x^m y^n$ represents the series then $a_{m,n+1}/a_{mn} = f(m, n)$ and $a_{m+1,n}/a_{mn} = g(m, n)$ where f and g are rational functions of m and n.

A further extension of this conception is toward the consideration of series of the form $z = \sum_0^\infty \sum_0^\infty a_{mn} x_m y_n$ in which the coefficients, a_{mn}, satisfy two linear difference equations

$$(1) \qquad \begin{aligned} \sum\sum P_{rs}(m, n)\, a_{m+r,n+s} &= 0, \\ \sum\sum Q_{rs}(m, n)\, a_{m+r,n+s} &= 0, \end{aligned}$$

P_{rs} and Q_{rs} indicating polynomials in m and n.[8]

Another suggestion put forward is that the definition of the hypergeometric function of two variables might well be extended by generalizing Pochhammer's integral in the form

$$(2) \quad z = \int_g^h (u-a_1)^{b_1-1}(u-a_2)^{b_2-1}\cdots(u-a_n)^{b_n-1}(u-x)^{\lambda-1}(u-y)^{\mu-1}\,du.$$

Where g and h represent any two of the quantities $a_1, a_2, \cdots, a_k, x, y, \infty$. It is the object of this note to show how the functions defined by this integral are included in the functions which may be expressed by the series whose coefficients satisfy the equations (1).

Let us first consider the function

$$(3) \qquad G = \prod_{i=1}^{i=k} (u-a_i)^{b_i} w^{\mu-\beta}(u-x)^{\lambda-k-1},$$

and suppose that a_1 and a_2 are of the same sign, and that $b_i > 0$ if $a_2 \geqq a_i \geqq a_1$.

[7] Appell, Kampé de Feriet, Fonctions hypergéométriques et hypersphériques—Polynomes d'Hermite, Paris, 1926. Also, Appell, Sur les fonctions hypergéométriques de plusieurs variables, les polynomes d'Hermite et autres fonctions sphériques dans l'hyperespace, Mémorial des Sciences Mathématiques, Paris, 1925. Each of these works includes a bibliography of the subject.

[8] Appell, Kampé de Feriet, loc. cit., p. 147. Also, Appell, loc. cit., p. 41.

From (3) we obtain by differentiation

$$\frac{dG}{du} = \left\{(\lambda-k-1)\left(\sum_{i=0}^{k+1} A_i u^{k+1-i}\right)(u-x)^{\lambda-k-2}\right.$$
$$\left. + \left[(\mu-\beta)\left(\sum_{i=1}^{i=k} A_i u^{k-i}\right) + \left(\sum_{i=0}^{i=k} B_i u^{k-i}\right)\right](u-x)^{\lambda-k-1}\right\} U; \tag{4}$$

where

$$\sum_{i=0}^{i=k} A_i u^{k-i} = \prod_{i=1}^{i=k}(u-a_i), \quad \sum_{i=0}^{i=k} B_i u^{k-i-1} = \left(\sum_{i=1}^{i=k}\frac{b_i}{u-a_i}\right)\prod_{i=1}^{i=k}(u-a_i),$$
$$A_0 = 1, \quad A_{k+1} = B_k = B_{k+1} = 0;$$

and

$$U = \prod_{i=1}^{i=k}(u-a_i)^{b_i-1} u^{\mu-\beta-1}.$$

Making use of the identity

$$\sum_{r=0}^{r=k+1} A_r u^{k+1-r} = \sum_{j=0}^{j=k+1}\frac{1}{(k+1-j)!}\sum_{i=0}^{i=j}\frac{(k+1-i)!}{(j-i)!} A_i x^{j-i}(u-x)^{k+1-j}$$

we may write

$$\frac{dG}{du} = \left\{(\lambda-k-1)\left(\sum_{j=0}^{j=k+1}\frac{1}{(k+1-j)!}\sum_{i=0}^{i=j}\frac{(k+1-i)!}{(j-i)!} A_i x^{j-i}(u-x)^{\lambda-1-j}\right)\right.$$
$$\left. + \sum_{j=0}^{j=k}\frac{1}{(k-j)!}\sum_{i=0}^{i=j}\frac{(k-i)!}{(j-i)!}\left((\mu-\beta)A_i + B_i\right)x^{j-i}(u-x)^{\lambda-1-j}\right\} U.$$

Now since $\int_{a_1}^{a_2}\frac{dG}{du}\,du = 0$ it follows that

$$(\lambda-k-1)\sum_{j=0}^{j=k+1}\frac{1}{(k+1-j)!}\sum_{i=0}^{i=j}\frac{(k+1-i)!}{(j-i)!}$$
$$\times A_i x^{j-i}\int_{a_1}^{a_2} U(u-x)^{\lambda-j-1}\,du$$
$$+\sum_{j=0}^{j=k}\frac{1}{(k-j)!}\sum_{i=0}^{i=k}\frac{(k-i)!}{(j-i)!}\left((\mu-\beta)A_i + B_i\right) \tag{5}$$
$$\times x^{j-i}\int_{a_1}^{a_2} U(u-x)^{\lambda-j-1}\,du = 0.$$

Then if we let

$$z = \int_{a_1}^{a_2}\prod_{i=1}^{i=k}(u-a_i)^{b_i-1}(u-x)^{\lambda-1}(u-y)^{\mu-1}\,du$$

it follows that

$$\left.\frac{\partial^{j+\beta} z}{\partial x^j\,\partial y^\beta}\right|_{y=0} = (-1)^{j+\beta}\frac{\Gamma(\lambda)}{\Gamma(\lambda-j)}\frac{\Gamma(\mu)}{\Gamma(\mu-\beta)}\int_{a_1}^{a_2} U(u-x)^{\lambda-1-j}\,du. \tag{6}$$

Also, it can be shown that when z is expressed by means of a double power series, $z = \sum_{r=0}^{r=\infty} \sum_{s=0}^{s=\infty} a_{rs}\, x^r y^s$, we have

$$(7) \qquad \frac{\partial^\alpha}{\partial x^\alpha}\left(x^r \frac{\partial^{p+q} z}{\partial x^p\, \partial y^q}\right)\Bigg|_{\substack{y=0\\x=0}} = \frac{\alpha!\,(p-r+\alpha)!\,q!}{(\alpha-r)!}\, a_{p-r+\alpha,\,q}.$$

We now express the integrals in (5) by means of the partial derivatives of z, making use of (6). Then we differentiate (5) α times with respect to x, and let x be zero. These operations and the relations (7) enable us to express equation (5) in the form

$$(\lambda-k-1)\sum_{j=0}^{j=k+1} \frac{\Gamma(\lambda-j)\,(-1)^j}{\Gamma(\lambda)\,(k+1-j)!} \sum_{i=0}^{i=j} \frac{(k+1-i)!}{(j-i)!}\, A_i\, \frac{\alpha!\,(i+\alpha)!\,\beta!}{(\alpha-j+i)!}\, a_{\alpha+i,\beta}$$
$$+\sum_{j=0}^{j=k} \frac{\Gamma(\lambda-j)\,(-1)^j}{\Gamma(\lambda)\,(k-j)!} \sum_{i=0}^{i=k} \frac{(k-i)!}{(j-i)!}\left((\mu-\beta)\,A_i + B_i\right) \frac{\alpha!\,(i+\alpha)!\,\beta!}{(\alpha-j+i)!}\, a_{\alpha+i,\beta} = 0.$$

Interchanging the order of summation in the preceeding equation it may be written

$$(8) \qquad \begin{aligned} &(\lambda-k-1)\sum_{i=0}^{i=k+1} (k+1-i)!\,(\alpha+i)! \\ &\qquad \times a_{\alpha+i,\beta}\, A_i \sum_{j=i}^{j=k+1} \frac{\Gamma(\lambda-j)\,(-1)^j}{(k+1-j)!\,(j-i)!\,(\alpha-j+i)!} \\ &\quad +\sum_{i=0}^{i=k} (k-i)!\,(\alpha+i)! \\ &\qquad \times a_{\alpha+i,\beta}\left((\mu-\beta)\,A_i + B_i\right) \sum_{j=i}^{j=k} \frac{\Gamma(\lambda-j)\,(-1)^j}{(k-j)!\,(j-i)!\,(\alpha-j+i)!} = 0. \end{aligned}$$

We shall now show how

$$(9) \qquad \begin{aligned} &\frac{1}{\Gamma(\lambda-k-1)} \sum_{j=i}^{j=k+1} \frac{\Gamma(\lambda-j)\,(-1)^j}{(k+1-j)!\;(j-i)!\;(\alpha-j+1)!} \\ &= \frac{(-1)^{k+1}}{\Gamma(\lambda-k-1)} \sum_{s=0}^{s=k+1-i} \frac{\Gamma(\lambda+s-k-1)\,(-1)^s}{s!\,(k+1-s-i)!\,(\alpha-k-1+s+i)!}, \end{aligned}$$

a polynomial of degree $k+1-i$ in λ, may be expressed as a product of linear factors.

If we let $\lambda = \alpha+k+2-p$ the summation in the second member of (9) becomes

$$(10) \qquad \sum_{s=0}^{s=k+1-i} \frac{\Gamma(\alpha+s-p+1)\,(-1)^s}{s!\,(k+1-s-i)!\,(\alpha-k-1+s+i)!}.$$

In order to factor this expression we consider the identity

$$(1-x)^{p-1} \equiv (1-x)^{k+1-i}\,(1-x)^{p-1-k-1+i}.$$

If α, p, k, i are integers such that $\alpha \geqq k+1-i > p-1$ then the coefficient of x^α on the left hand side is zero. Setting the coefficient on the right hand side equal to zero we find

$$0 = \frac{(k+1-i)!}{(k+1-i-p)!} \sum_{s=0}^{s=k+1-i} \frac{(\alpha+s-p)!\,(-1)^s}{(\alpha+s-k-1+i)!\,(k+1-i-s)!\,s!}.$$

Since this is true for $p = 1, 2, \cdots, k+1-i$ the expression (9) is of the form $R\,\dfrac{\Gamma(\alpha+k-\lambda+2)}{\Gamma(\alpha+i-\lambda+1)}$, where R is independent of λ.

To determine R we equate the coefficient of λ^{k+1-i} in (9) to the corresponding coefficient in the expression which we have just obtained. The result is

$$\frac{(-1)^i}{(k+1-i)!\,\alpha!} = (-1)^{k+1-i} R,$$

and (9) reduces to

$$(-1)^{k+1} \frac{\Gamma(\alpha+k-\lambda+2)}{\Gamma(\alpha+i-\lambda+1)\,(k+1-i)!\,\alpha!}.$$

Introducing this result in (8) and employing a similar one for the second term of the left member of that equation we obtain from it the difference equation between the successive coefficients in the power series development. The difference equation is

$$(11) \quad \sum_{i=0}^{i=k+1} [(\alpha+k-\lambda+1-\mu+\beta)\,A_i - B_i] \frac{(\alpha+i)!\,\Gamma(\alpha+k-\lambda+1)\,a_{\alpha+i,\beta}}{\alpha!\,\Gamma(\alpha+i-\lambda+1)} = 0.$$

Since z satisfies Euler's equation in the form

$$(12) \qquad (x-y)\frac{\partial^2 z}{\partial x\,\partial y} + (\mu-1)\frac{\partial z}{\partial x} - (\lambda-1)\frac{\partial z}{\partial y} = 0,$$

we can obtain another difference equation between the coefficients in the power series $z = \sum_{r=0}^{r=\infty} \sum_{s=0}^{s=\infty} a_{r,s}\,x^r\,y^s$. It is

$$(13) \quad (\beta+1)(\alpha-\lambda+1)\,a_{\alpha,\beta+1} - (\alpha+1)(\beta-\mu+1)\,a_{\alpha+1,\beta} = 0.$$

If k coefficients, $a_{m,n}$, $m+n = 0, 1, 2, \cdots, k-1$, are chosen arbitrarily the two difference equations (11) and (13) enable us to determine uniquely any other coefficient.

SUR UNE PROPRIÉTÉ MÉTRIQUE DE LA DIMENSION.[1]

PAR L. PONTRJAGIN ET L. SCHNIRELMANN.

Le but de cette note est d'obtenir la dimension d'un espace métrique compact à l'aide d'un procédé de recouvrement par des « étalons de mesure » dont le nombre permet de définir la dimension. Cette méthode est analogue à la méthode de mesure employée en géométrie élémentaire.

Dans les définitions de la dimension on se sert toujours de notions purement topologiques et on laisse de côté les propriétés métriques. Comme on va le voir, il existe cependant un lien fort frappant entre les propriétés métriques d'un ensemble compact et sa dimension.

Nous emploierons la forme suivante de la définition de dimension (au sens bien connu de Brouwer, Urysohn et Menger):

Un ensemble compact F a la dimension n, si n est le plus petit entier tel qu'il existe (ε étant un nombre arbitraire positif) un système fini d'ensembles fermés recouvrant F, de diamètres ne surpassant pas ε et dont toutes les intersections de $n+2$ sont vides.

Soit donné un espace métrique compact F. Nous le recouvrons par un système d'un nombre fini d'ensembles fermés de diamètres ne surpassant pas ε. Ceci est possible d'après le théorème de Heine-Borel. Nous pouvons définir le nombre minimum $N(\varepsilon)$ de tels ensembles nécessaire pour recouvrir F. $N(\varepsilon)$ prend des valeurs entières positives pour toutes les valeurs positives de ε et croît indéfiniment quand ε tend vers 0 (à moins que l'ensemble F ne contienne qu'un nombre fini de points). Cette fonction dépend évidemment de la métrique de l'espace F. Nous l'appellerons fonction de volume de l'espace F relative à la métrique donnée et nous la désignerons par $N_F(\varepsilon)$.

Soit r un nombre réel quelconque; nous dirons que r appartienne à la première classe s'il existe un nombre positif c tel qu'on ait

$$N_F(\varepsilon) \geqq \frac{c}{\varepsilon^r} \quad \text{(quel que soit } \varepsilon > 0\text{)};$$

dans le cas contraire r appartiendra à la seconde classe. La coupure de Dedekind ainsi obtenue détermine un nombre non négatif k ($\leqq +\infty$) que nous appelons *ordre métrique de l'espace F*. On voit d'ailleurs sans peine que

$$k = \liminf \left(- \frac{\log N_F(\varepsilon)}{\log \varepsilon} \right).$$

[1] Received July 11, 1931.

Pour les diverses métriques de F nous obtenons les diverses fonctions de volume et les divers ordres métriques. Par exemple, un arc simple de Jordan, étant métrisé de la façon euclidienne, possède comme fonction de volume (asymptotiquement) $\frac{c}{\varepsilon}$, où c est une constante. L'ordre métrique d'un tel arc est donc égale à *un*. D'autre part, nous pouvons définir une autre métrique d'un arc simple de la façon suivante: prenons un ensemble parfait partout discontinu dans l'espace euclidien à k dimensions R^k dont la mesure est positive. Par un tel ensemble on peut mener un arc simple. Cet arc possède automatiquement la métrique de l'espace euclidien R^k, sa fonction de volume est asymptotiquement $\frac{c}{\varepsilon^k}$ et son ordre métrique est égale à k.

Ainsi l'ordre métrique n'est point un invariant de F. Or, on obtient évidemment un invariant topologique de l'espace F en considérant la borne inférieure des ordres métriques correspondant à diverses métriques qu'on peut imposer à F sans altérer les propriétés topologiques de cet espace. Le but principal du présent travail est de déterminer l'invariant topologique en question et *de montrer qu'il est égale à la dimension Brouwerienne de l'espace* F. On a donc le

THÉORÈME FONDAMENTAL. *La borne inférieure de l'ordre métrique*

$$k = \liminf\left(-\log\frac{N_F(\varepsilon)}{\log\varepsilon}\right)$$

pour toutes les métriques de l'espace compact F *est égale à sa dimension.*

COROLLAIRE. *Si la borne inférieure en question est finie, elle est nécessairement un nombre entier.*

Démonstration.

1. Soit K un cube euclidien à r dimensions. Il est évident que, dans sa métrique ordinaire, on a (pour $\varepsilon>0$ assez petit et des constantes c et c' convenablement choisies)

$$\frac{c'}{\varepsilon^r} > N_K(\varepsilon) > \frac{c}{\varepsilon^r}.$$

2. Nous dirons qu'une transformation φ d'un espace métrique F en un espace métrique F' est une transformation sans dilatation, si la distance de chaque paire de points de F ne croît pas en vertu de cette transformation, c'est à dire si

$$\varrho[\varphi(a), \varphi(b)] \leqq \varrho(a, b).^2$$

Toute transformation sans dilatation est évidemment continue.

[2] Les transformations de cette sorte ont été étudiées pour la première fois par M. Kolmogoroff dans ses recherches sur la théorie de la mesure.

Il est facile de voir que s'il est possible de transformer sans dilatation un espace métrique F en un espace métrique F', on a

$$N_{F'}(\varepsilon) \leqq N_F(\varepsilon).$$

3. Montrons que pour un espace compact de dimension $\geqq r$ finie ou infinie, on a:

$$N_F(\varepsilon) > \frac{c}{\varepsilon^r}$$

(ε suffisamment petit, c constante positive, dépendante de F.)

On démontre d'abord la proposition suivante:

THÉORÈME AUXILIAIRE. *Un espace métrique compact F, de dimension $\geqq r$ finie ou infinie, peut être transformé sans dilatation en un sous-ensemble r-dimensionnel d'un espace Euclidien à r dimensions.*

Démonstration du théorème auxiliaire. Soit

$$\text{(I)} \qquad a_1, a_2, \cdots, a_n$$

un groupe arbitraire fini de points sur F. A chaque point x de F correspond le système des nombres

$$\text{(II)} \qquad \varrho_1, \varrho_2, \cdots, \varrho_n,$$

où $\varrho_1, \varrho_2, \cdots, \varrho_n$ sont les distances $\varrho(a_i, x) = \varrho_i$ entre x et les points du groupe choisi. Faisons correspondre à x un point $\varphi'(x)$ d'un espace euclidien à n dimensions aux coordonnées cartésiennes $x_1, x_2, \cdots, x_n$, où

$$\text{(III)} \qquad x_i = \frac{\varrho_i}{\sqrt{n}}.$$

Démontrons que la transformation φ' ainsi définie est sans dilatation. En effet, p' et p'' étant deux points arbitraires sur F et $r_1', r_2', \cdots, r_n'$; $r_1'', r_2'', \cdots, r_n''$ les coordonnées correspondantes de leurs transformés, nous avons:

$$\varrho[\varphi'(p'), \varphi'(p'')] = \sqrt{(r_1''-r_1')^2+(r_2''-r_2')^2+\cdots+(r_n''-r_n')^2}.$$

Soit j un indice tel que

$$|r_j''-r_j'| \geqq |r_i''-r_i'| \qquad (i = 1, 2, \cdots, n).$$

Nous avons:

$$\varrho[\varphi'(p'), \varphi'(p'')] \leqq (r_j''-r_j')\sqrt{n} = \varrho_j''-\varrho_j'.$$

Mais en vertu de l'axiome du triangle, on a:

$$|\varrho_j''-\varrho_j'| \leqq \varrho(p', p'').$$

Par suite

$$\varrho[\varphi'(p'), \varphi'(p'')] \leqq \varrho(p', p'').^3$$

[3] Cette partie de la démonstration n'est qu'une légère modification d'un raisonnement d'Urysohn (Comptes rendus Paris, 178 (1924), p. 65).

Démontrons maintenant que pour un choix convenable du système (I), la dimension de $F' = \varphi'(F)$ ne sera pas plus petite que r. Prenons d'abord (I) de telle façon qu'à chaque point $x \subset F$ corresponde un point a_j du système (I), tel que $\varrho(a_j, x) < \frac{d}{2}$, où d est le r^{me} coëfficient d'aplatissement[3a] de l'ensemble F (ce nombre est nécessairement positif puisque la dimension de F n'est pas inférieure à r).[4]

Démontrons d'abord l'existence d'un nombre positif δ vérifiant la condition: Si $\varrho[\varphi'(p), \varphi'(q)] < \delta$ on a $\varrho(p, q) < d$. En effet, supposons au contraire qu'il existe sur F deux suites infinies convergente de points

(IV) $$p_1, p_2, \cdots, p_m, \cdots$$

et

(V) $$q_1, q_2, \cdots, q_m, \cdots$$

telles que

$$\varrho(p_m, q_m) \geqq d \quad \text{avec} \quad \varrho[\varphi'(p_m), \varphi'(q_m)] \to 0 \quad (m \to \infty).$$

En désignant par p et q deux points limites de p_m et q_m, nous avons:

$$\varrho[\varphi'(p), \varphi'(q)] = 0, \qquad \varrho(p, q) \geqq d,$$

c'est à dire

(VI) $$\begin{cases} \varphi'(p) = \varphi'(q) \\ \varrho(p, q) \geqq d. \end{cases}$$

Il existe, par hypothèse, un point a_j du système (I) pour lequel $\varrho(a_j, p) < \frac{d}{2}$. Pour ce point $\varrho(a_j, q) > \frac{d}{2}$ (parce que dans le cas contraire on aurait

$$\varrho(p, q) \leqq \varrho(a_j, p) + \varrho(a_j, q) < d).$$

Par conséquent les j^{mes} coordonnées cartésiennes des points $\varphi'(p)$ et $\varphi'(q)$ sont distinctes, ce qui contredit la condition (VI).

Supposons que la dimension de F' soit moindre que r. Il existe alors une couverture de F' par un système d'ensembles fermés $f_1', f_2', \cdots, f_k'$, de diamètres moindres que δ et l'ordre de cette couverture (c'est à dire le nombre maximum d'ensembles de la couverture qui peuvent passer par un même point) ne surpasse pas r.

En désignant par f_i l'ensemble de tous les points de F dont l'image appartient à f_i', nous obtenons une couverture de l'ensemble F par les ensembles f_i de diamètre $< d$; l'ordre de cette couverture ne surpasse pas r.

[3a] Au sens d'Urysohn, Fund. Math. 8 (1926), p. 353.

[4] Un tel choix du système (I) est toujours possible, d étant positif, et F compact.

Or cette dernière conclusion est impossible parce que le nombre $r^{\text{ième}}$ d'Urysohn de F est égal à d.

Supposons cependant que le système (I) soit choisi de telle sorte, que la dimension de F' n'est pas inférieure à r et dans cette hypothèse faisons la construction suivante: soit $R_1, R_2, \cdots, R_e$ l'ensemble de tous les plans de coordonnées à r dimensions de R^n. Désignons par $\varphi_i(F')$ la projection orthogonale de F' sur R_i; nous écrirons, pour abréger, F_i^* pour l'ensemble $\varphi_i(F')$. Nous allons démontrer qu'un au moins des F_i^* a la dimension r.

Soit Q un cubillage de Lebesgue[6] de R^n, les faces duquel sont parallèles aux plans des coordonnées. Soit

(VII) $$p_1, p_2, \cdots, p_t, \cdots$$

l'ensemble de toutes les intersections non vides de $r+1$ cubes du cubillage Q. Le système (VII) ne contient qu'un nombre fini d'éléments qui coupent F'. Soit p_s un de ces éléments. Démontrons que si toutes les dimensions de toutes les projections F_i^* sont moindres que r, il existe un écart parallèle arbitraiment petit ψ de l'espace R^n tel que $\psi(p_s)$ ne coupe pas F'.

Soit en effet R^{n-r} l'espace $n-r$-dimensional qui porte p_s. A cause du parallélisme du R^{n-r} à l'un des plans de coordonnées, R^{n-r} est perpendiculaire à un des plans R_i, par exemple à R_t, et coupe R_t en un point. D'après la supposition que la dimension de F_t^* est moindre que r, nous concluons que F_t^* ne contient aucun point intérieur de R_t, c'est à dire qu'on peut choisir un écart parallèle ψ arbitraiment petit de telle sorte que les intersections de $\psi(R^{n-r})$ avec F_t^* et par conséquent de $\psi(R^{n-r})$ avec F' et $\psi(p_s)$ avec F', soient vides.

On peut supposer l'écart ψ si petit, que si l'intersection de p_i avec F' est vide, l'intersection de $\psi(p_i)$ avec F' est vide aussi. En répétant l'opération d'écart parallèle un nombre fini de fois, nous obtenons un écart parallèle $\overline{\psi}$, tel que toutes les intersections de $\overline{\psi}(p_i)$ avec F' sont vides ($i = 1, 2, \cdots$).

Désignons par

(VIII) $$q_1, q_2, \cdots, q_s$$

l'ensemble de tous les cubes du cubillage $\overline{\psi}(Q)$, dont les intersections avec F' ne sont pas vides et par $\overline{f_i}$ l'intersection de q_i avec F'. Les intersections de tous les produits de $r+1$ cubes du système (VIII) ne coupent pas F'. Le système $\overline{f_1}, \overline{f_2}, \cdots, \overline{f_s}$ forme donc une couverture de F' d'ordre ne surpassant pas r.

Mais le cubillage Q (et par conséquent $\overline{\psi}(Q)$) étant de petitesse arbitraire, la dimension de F' est nécessairement plus petite que r — ce qui est impossible.

[6] Fund. Math. 2 (1921), p. 266.

Nous pouvons donc conclure qu'une au moins parmi les projections F_i^* de l'ensemble F', par exemple F_j^*, est de dimension r. Comme la projection est une transformation sans dilatation, la transformation $\varphi(F) = \varphi_j[\varphi'(F)]$ en est aussi une et elle transforme F en un ensemble F_j^* r-dimensionnel situé dans l'espace euclidien à r dimensions. Le théorème auxiliaire est donc démontré.

Le ensemble que nous venons d'obtenir contient nécessairement un point intérieur de R^r et l'on déduit facilement de là que, ε étant suffisamment petit,

$$N_F(\varepsilon) \geqq N_{F_j^*}(\varepsilon) \geqq \frac{c}{\varepsilon^r}. \tag{1}$$

4. Démontrons que r est précisemment la borne inférieure de tous les exposants

$$\liminf\left[-\frac{\log N_F(\varepsilon)}{\log \varepsilon}\right].$$

Dans ce but il suffit d'obtenir une métrique de F dans laquelle $N_F(\varepsilon)$ devient moindre que $\frac{c}{\varepsilon^{r+\delta(\varepsilon)}}$, ou $\delta(\varepsilon)\to 0$ si $\varepsilon\to 0$.

Nous partons du l'approximation des ensembles dans l'espace euclidien.[7]

Étant donné l'espace métrique compact F de dimension r, il est possible de trouver dans l'espace euclidien R^{2r+1} à $2r+1$ dimensions, une suite $C_1, C_2, \cdots, C_n \cdots$ de complexes polyédraux à r dimensions, possédant les propriétés suivantes:

1. la suite a une limite topologique F' qui est homéomorphe à F.
2. on peut la construire de telle manière que si les n premiers termes $C_1, C_2, \cdots, C_n$ sont déjà construits, on peut alors placer le $(n+1)^{\text{me}}$ C_{n+1} dans un voisinage arbitraire de C_n.

Nous pouvons supposer que tous les complexes de la suite $C_1, C_2, \cdots$ soient situés dans une partie finie G de l'espace R^{2r+1}. Les complexes C_n étant polyédraux et la structure combinatoire de C_i étant définie d'avance, il est possible de choisir pour chaque indice i et chaque nombre positif δ_i un nombre positif ε_i tel que $N_{C_i}(\varepsilon) < \left(\frac{1}{\varepsilon}\right)^{r+\delta_i}$ (pour $\varepsilon \leqq \varepsilon_i$). On peut encore supposer que $\varepsilon_i = \tau^{p_i}$, où $0<\tau<1$ et les p_i sont des entiers positifs avec $p_{i+1} > p_i$; $\delta_1, \delta_2, \cdots$ étant une suite arbitraire de nombres positifs tendant vers 0.

Supposons que C_1 ait une position déterminée dans G. Désignons par φ_j, $j = p_1, p_1+1, \cdots, p_2-1$ une couverture du complexe C_1 par des

[7] Voir pour la démonstration un article de L. Pontrjagin et M^elle Tolstowa à paraître dans le tome 105 des „Mathematische Annalen". Voir aussi Lefschetz, Annals of Mathematics 1931, vol. 32, no. 3.

ensembles de diamètres moindres que τ^j, telle que le nombre n'en surpasse pas $\left(\frac{1}{\tau^j}\right)^{r+\delta_1}$. En élargissant tous les ensembles de la couverture de telle façon que les diamètres des ensembles élargis soient aussi moindres que τ^j, nous obtenons une couverture φ'_j d'un voisinage G'_j du complexe C_1. Le nombre des éléments de la couverture φ'_j est égal à celui de φ_j, c'est à dire ne surpasse pas $\left(\frac{1}{\tau^j}\right)^{r+\delta_1}$.

En désignant par G_1 l'intersection des voisinages $G'_{p_1}, G'_{p_1+1}, \cdots, G'_{p_2-1}$ nous obtenons:

$$N_{\bar{G}_1}(\varepsilon) \leqq \left(\frac{1}{\varepsilon}\right)^{r+\delta_1} \text{ (pour } \varepsilon = \tau^j;\ j = p_1,\ p_1+1,\ \cdots,\ p_2-1).$$

Construisons le complexe C_2 à l'intérieur du voisinage G_1. En répétant les constructions précédentes, nous obtenons une suite de complexes $C_1, C_2, \cdots, C_i, \cdots$ et de leurs voisinages $G_1, G_2, \cdots, G_i, \cdots$ qui satisfont aux conditions suivantes:

1. la clôture $\bar{G}_i$ de G_i est située dans G_{i-1};
2. pour chaque indice j, $p_{i-1} \leqq j \leqq p_i - 1$, on a:

$$N_{\bar{G}_{i-1}}(\tau^j) \leqq \left(\frac{1}{\tau^j}\right)^{r+\delta_{i-1}}.$$

Mais la limite topologique F^r de la suite $C_1, C_2, \cdots, C_n, \cdots$ est située dans l'intersection de tous les $\bar{G}_1, \bar{G}_2, \cdots, \bar{G}_n, \cdots$.

Par suite:

$$N_{F^r}(\tau^j) \leqq \left(\frac{1}{\tau^j}\right)^{r+\delta'(j)} \quad (j>0,\ \delta'(j)>0 \text{ et } \delta'(j) \text{ tend vers 0, quand } j \text{ croît indéfinément}).$$

Soit $\varepsilon < \tau$ un nombre positif arbitraire. Nous pouvons le mettre sous la forme

$$\varepsilon = s\tau^j, \quad \tau < s \leqq 1$$

et nous obtenons:

$$N_{F^r}(s\tau^j) \leqq N_{F^r}(\tau^{j+1}) < \left(\frac{1}{\tau^{j+1}}\right)^{r+\delta'(j+1)} \leqq \left(\frac{1}{s\tau^j}\right)^{r+\delta'(j+1)+\frac{\log s - \log \tau}{-\log(s\tau^j)}}$$

c'est à dire $N_F(\varepsilon) < \left(\frac{1}{\varepsilon}\right)^{r+\delta(\varepsilon)}$.

Notre théorème fondamental sur la dimension est donc complètement démontré.

Moscou,
mai 1931.

ÜBER STETIGE ALGEBRAISCHE KÖRPER.*

VON L. PONTRJAGIN IN MOSKAU.

§ 1.

Der Zweck dieser Arbeit ist, eine topologische Charakterisierung der Körper der reellen und der komplexen Zahlen zu geben (Satz II). Wenn wir auf die Kommutativitätsbedingung, die in dieser Charakterisierung eine wesentliche Rolle spielt, verzichten, erhalten wir noch den Quaternionenkörper (Satz I).

Die Fragestellung, die zu dieser Arbeit geführt hat, sowie zahlreiche Anregungen bei der Beweisführung und bei der Redaktion der vorliegenden Abhandlung verdanke ich Herrn A. Kolmogoroff.

§ 2.

Definition I. Eine Menge K von Elementen $a, b, c, \cdots$ bildet einen *stetigen Körper*, wenn folgende Bedingungen erfüllt sind:

1. K ist ein topologischer Raum,[1]
2. K ist ein algebraischer, im allgemeinen nicht kommutativer Körper,
3. die Funktionen $f_1(a, b) = a+b$, $f_2(a, b) = ab$, $f_3(a) = -a$, $f_4(a) = a^{-1}$ sind stetig (überall, wo sie definiert sind, d. h. f_1, f_2 und f_3 im ganzen Raum und f_4 überall mit Ausnahme des Nullpunktes).

Im folgenden wird unter einem *Körper* stets ein stetiger Körper verstanden.

Definition II. Zwei stetige Körper heißen isomorph, wenn sie aufeinander eineindeutig abgebildet werden können, und zwar so, daß dabei sowohl die algebraischen als auch die topologischen Relationen in beiden Körpern einander entsprechen (d. h. daß die Abbildung gleichzeitig einen Isomorphismus der Körper und einen Homöomorphismus der Räume darstellt).

* Received May 17, 1931.

[1] Eine Menge R von irgendwelchen Elementen soll ein topologischer Raum heißen, (vgl. Alexandroff, Math. Ann. 96, S. 555), wenn für jede Teilmenge M von R die abgeschlossene Hülle $\overline{M}$ definiert ist, und zwar so, daß folgende Bedingungen erfüllt sind:

I. Eine aus einem einzigen Punkt bestehende Menge ist mit ihrer abgeschlossenen Hülle identisch.

II. $\overline{A+B} = \overline{A} + \overline{B}$.

III. $\overline{(\overline{A})} = \overline{A}$.

Mengen, die mit ihrer abgeschlossenen Hülle zusammenfallen, heißen abgeschlossen, ihre Komplementärmengen offen. Jede offene Menge, die einen gegebenen Punkt enthält, heißt eine *Umgebung* dieses Punktes.

SATZ I. *Jeder stetige lokalbikompakte,*[2] *zusammenhängende Körper K ist dem Körper der reellen, dem Körper der komplexen Zahlen oder dem Quaternionenkörper isomorph.*

SATZ II. *Jeder stetige lokalbikompakte, zusammenhängende kommutative Körper K ist entweder dem Körper der reellen oder dem Körper der komplexen Zahlen isomorph.*

In der Formulierung der Sätze I und II kann die Bikompaktheit durch die Bedingung ersetzt werden, daß K ein Hausdorffscher lokalkompakter topologischer Raum mit erstem Abzählbarkeitsaxiom ist:

SATZ Ia. *Jeder stetige lokalkompakte, zusammenhängende Körper K, in dem das erste Abzählbarkeitsaxiom und das Hausdorffsche Trennbarkeitsaxiom erfüllt sind, ist mit dem Körper der reellen Zahlen, dem Körper der komplexen Zahlen oder dem Quaternionenkörper isomorph.*

SATZ IIa. *Unter den Bedingungen des Satzes* Ia *mit Hinzufügung der Kommutativität von K ist K entweder dem Körper der reellen Zahlen oder dem Körper der komplexen Zahlen isomorph.*

Wir werden beweisen, daß die Bedingungen von Satz Ia aus den Bedingungen von Satz I folgen, worauf wir nur noch den Satz Ia beweisen müssen: II und IIa folgen ja unmittelbar aus I und Ia infolge der Nichtkommutativität des Quaternionenkörpers.

§ 3.

In jedem stetigen Körper K gilt folgender Satz, der keines Beweises bedarf:

HILFSSATZ I. *Wenn $a \neq 0$ ist, so ist die Abbildung*

$$x' = ax + b$$

von K auf sich eine topologische (d. h. eineindeutige und beiderseits stetige) *Abbildung.*

Dieser Satz zeigt, daß K topologisch *homogen* ist, d. h. daß man mittels einer topologischen Abbildung des ganzen Körpers auf sich selbst jeden Punkt a in jeden Punkt b überführen kann: es genügt in der Tat, die Abbildung $x' = x - a + b$ heranzuziehen.

§ 4.

In diesem Paragraphen setzen wir voraus, daß der Körper K lokalbikompakt und zusammenhängend ist.

---

[2] Wegen der Kompaktheit bzw. Kompaktheit im kleinen, sowie der weiteren hier gebrauchten topologischen Begriffe (wie Trennungsaxiome, insb. Regularität, Abzählbarkeitsaxiome u. dgl.), siehe Alexandroff, Math. Ann. 96, S. 555, sowie Alexandroff und Urysohn, Math. Ann. 92.

HILFSSATZ II. *Es existiert eine abzählbare Menge* $a_1, a_2, \cdots, a_r, \cdots$ (1) *von Elementen von* K, *die* 0 *zum Häufungspunkt hat.*

Beweis. Daraus, daß K zusammenhängend ist, folgt, daß alle seine Punkte Häufungspunkte sind (K enthält ja mindestens zwei Elemente: Null und Eins). Sei G eine offene Menge, deren abgeschlossene Hülle $\overline{G}$ kompakt ist. Da alle Elemente von K Häufungspunkte sind, enthält $\overline{G}$ unendlich viele Elemente. Sei $b_1, b_2, \cdots, b_r, \cdots$ (2) eine abzählbar unendliche Menge von Elementen aus $\overline{G}$ und b ein Häufungspunkt dieser Menge. Man sieht leicht, daß die Menge $b_1 - b, b_2 - b, \cdots, b_r - b, \cdots$ Null zum Häufungspunkt hat.

HILFSSATZ III. *Sei* F *eine bikompakte Menge*[3] *und*

$$a_1, a_2, \cdots, a_r, \cdots \tag{1}$$

eine Folge von Elementen, die Null zum Häufungspunkt hat und G *eine Umgebung der Null, so existiert ein* n, *so daß* $F_n = Fa_n \subset G$.[3a]

Beweis. Für jeden Punkt a gibt es stets eine Umgebung $G(a)$ von a und eine Umgebung $G'(a)$ von Null, so daß, wenn $x \subset G(a)$ und $y \subset G'(a)$ ist, xy in G liegt. Wenn nun a die ganze Menge F durchläuft, so überdecken die Gebiete $G(a)$ die ganze Menge F; infolge der Bikompaktheit von F kann man aus diesem Überdeckungssystem ein endliches Teilsystem $G(a^{(1)}), G(a^{(2)}), \cdots, G(a^{(k)})$ herausgreifen, welches ebenfalls F überdeckt; bezeichnen wir nun mit G' den Durchschnitt $G'(a^{(1)}), G'(a^{(2)}), \cdots, G'(a^{(k)})$. Man sieht leicht, daß das Produkt eines beliebigen Elementes aus F mit einem beliebigen Element aus G' in G liegt. Da aber G' eine Umgebung der Null ist, so gibt es unter den Elementen der Folge (1) sicher ein Element $a_n \subset G'$; daraus folgt aber $F_n \subset G$, w. z. b. w.

HILFSSATZ IV. K *ist ein regulärer Hausdorffscher Raum mit erstem Abzählbarkeitsaxiom.*

Beweis. Es sei U eine Umgebung des Nullpunktes mit bikompakter abgeschlossener Hülle $\overline{U}$ und b_n eine Folge, die Null zum Häufungspunkt hat. Bei beliebiger Wahl der Umgebung V des Nullpunktes ist bei entsprechend gewähltem n $\overline{U} \cdot b_n \subset V$, woraus folgt, daß die offenen Mengen $U \cdot b_n$ ein der Regularitätsbedingung genügendes abzählbares Umgebungssystem des Nullpunktes bilden.

Wegen der Homogenität von K (Hilfssatz I) existiert ein solches Umgebungssystem auch für jeden Punkt x. Daraus folgt bekanntlich auch das Hausdorffsche Trennungsaxiom (denn, wenn x und y zwei Punkte sind,

[3] d. h., daß F als Raum betrachtet bikompakt ist.

[3a] Wenn X eine Teilmenge von K und y ein Punkt von K ist, so bedeutet Xy die Menge aller Punkte xy, wobei x zu X gehört.

$V(x)$ zu y fremd, $\overline{U}(x)$ in $V(x)$ enthalten ist, so bilden $U(x)$ und das Komplement $K - \overline{U}(x)$ zwei disjunkte Umgebungen von x und y).

Auf diese Weise sehen wir, daß in K die Voraussetzungen des Satzes Ia erfüllt sind. Wir gehen also zum Beweis dieses Satzes über.

§ 5.

Wir setzen voraus, daß K lokalkompakt und zusammenhängend ist und dem ersten Abzählbarkeitsaxiom und dem Hausdorffschen Trennungsaxiom genügt. Ist x Häufungspunkt einer Menge E, so können wir aus E eine Folge herausgreifen, die gegen x *konvergiert.*[4] Dies ermöglicht uns, die ganze Untersuchung mit Hilfe des Begriffes der Konvergenz einer Folge zu führen.

HILFSSATZ V. *Wenn die Folge b_n gegen Null konvergiert, so ist die Folge b_n^{-1} divergent, d. h. sie besitzt keinen einzigen Häufungspunkt.*

Wenn nämlich c ein Häufungspunkt der Folge b_n^{-1} wäre, so hätte die Folge $x_n = b_n \cdot b_n^{-1} = 1$ den Punkt $0 \cdot c = 0$ zum Häufungspunkt, was offenbar unmöglich ist.

HILFSSATZ VI. *Zu jedem Punkt a gibt es eine gegen diesen Punkt konvergierende Folge von a verschiedenen Punkten.*

Beweis. Dies folgt daraus, daß a kein isolierter Punkt ist, und aus dem ersten Abzählbarkeitsaxiom.

HILFSSATZ VII. *K ist nicht kompakt.*

Es gibt nämlich nach Hilfssatz VI eine Folge b_n mit $b_n \neq 0$, $\lim b_n = 0$. Die Folge b_n^{-1} ist dann nach Hilfssatz V divergent.

HILFSSATZ VIII. *Wenn F kompakt,*[5] *V eine Umgebung des Nullpunktes und $\lim b_n = 0$ ist, so ist für hinreichend großes n $F \cdot b_n \subset V$.*

Man nehme das Gegenteil an. Dann gäbe es unendlich viele Produkte $x_n = f_n b_{k_n}$ ($f_n \subset F$), die zu V fremd sind. Nun haben aber die f_n einen Häufungspunkt f, während die b_{k_n} gegen 0 konvergieren. Die x_n haben daher den Häufungspunkt $f \cdot 0 = 0$, was unmöglich ist, da sie außerhalb V liegen.

HILFSSATZ IX. *Es seien U und V zwei Umgebungen des Nullpunktes mit kompakten abgeschlossenen Hüllen. Wenn a_n eine divergente Folge ist, so hat für hinreichend großes n die Begrenzung U' von U mit $\overline{V} \cdot a_n$ notwendig gemeinsame Punkte.*

Beweis. Wir betrachten den Durchschnitt $D(\overline{U}, \overline{V}a_n)$. Diese Menge ist kompakt und enthält den Nullpunkt. Da K selbst nicht kompakt ist, bildet sie eine echte Teilmenge von K. Da ferner K zusammenhängend

[4] Eine Folge $x_1, x_2, \cdots, x_n, \cdots$ konvergiert gegen x, wenn jede Umgebung von x alle bis auf endlich viele dieser Punkte enthält.

[5] "Kompakt" bedeutet *in sich* kompakt (d. h. kompakt und abgeschlossen).

ist, ist der Rand von $D(\bar{U}, \bar{V}a_n)$, d. h. die Menge $D(U', \bar{V}a_n) + D(\bar{U}, V'a_n)$, gewiß nicht leer. Wenn für unendlich viele n der erste Summand leer wäre, so wäre für dieselben n die Menge $D(\bar{U}, V'a_n)$ von 0 verschieden. Es würden m. a. W. die Folgen $u_k \subset \bar{U}$, $v_k \subset V'$ und eine Teilfolge a_{n_k} von a_n existieren, so daß $u_k = v_k a_{n_k}$ ist. Offenbar kann man überdies voraussetzen, daß u_k und v_k konvergente Punktfolgen mit

$$\lim u_k = u \subset \bar{U} \quad \text{und} \quad \lim v_k = v \subset V'$$

sind. Dabei ist v wegen der letzten Relation gewiß von 0 verschieden (V' ist ja die Grenze von V). Folglich ist $\lim a_{n_k} = \lim v_k^{-1} u_k = v^{-1} u$, was der Divergenz der Folge a_n widerspricht.

HILFSSATZ X. *Wenn die Folge a_n divergiert, so konvergiert die Folge a_n^{-1} gegen* 0.

Es genügt zu beweisen, daß, falls a_n divergiert, die Folge a_n^{-1} Null zum Häufungspunkt hat: daraus folgt nämlich, daß auch jede Teilfolge von a_n^{-1} den Nullpunkt zum Häufungspunkt hat, woraus sich vermöge des ersten Abzählbarkeitsaxioms ergibt, daß die ganze Folge a_n gegen Null konvergiert.

Es seien U und V zwei Umgebungen des Nullpunktes mit kompakter abgeschlossener Hülle (vgl. den vorigen Hilfssatz). Für alle hinreichend großen n ist $D(U', \bar{V}a_n)$ nicht leer, so daß man die Punkte v_n in $\bar{V}$ so wählen kann, daß für alle Punkte $u_n = v_n a_n$ zu U' gehören. Man kann Teilfolgen u_{n_k}, v_{n_k} auswählen, so daß u_{n_k} gegen $u \subset U'$ und v_{n_k} gegen $v \subset \bar{V}$ konvergieren, wobei $u \neq 0$ ist. Somit müßte $a_{n_k}^{-1} = u_{n_k}^{-1} v_{n_k}$ gegen $u^{-1} v$ konvergieren; da aber a_{n_k} divergent ist, muß $u^{-1} v$ der Nullpunkt sein, w. z. b. w.

HILFSSATZ XI. *Wenn a_n divergiert, F kompakt und V eine Umgebung des Nullpunkts ist, so ist für hinreichend großes n $F \subset Va_n$.*

In der Tat: a_n^{-1} konvergiert wegen Hilfssatz X gegen Null und daher ist für alle hinreichend großen n $Fa_n^{-1} \subset V$, woraus der Satz folgt.

HILFSSATZ XII. *Wenn F kompakt und V offen ist, p_n gegen p konvergiert und Fp in V enthalten ist, so gilt für hinreichend großes n: $Fp_n \subset V$.*

Sonst gäbe es eine Folge $f_k \subset F$, so daß für unendlich viele n, etwa n_k die Punkte $x_k = f_k p_{n_k}$ außerhalb V liegen würden. Da die p_{n_k} gegen p konvergieren und die f_{n_k} einen Häufungspunkt $f \subset F$ besitzen, so müssen die x_k den Punkt $fp \subset V$ als Häufungspunkt haben, was der Definition der x_k widerspricht.

HILFSSATZ XIII. *Wenn F beliebig, U abgeschlossen und $\lim p_n = p$ ist, so folgt aus $Fp_n \subset U$, daß auch $Fp \subset U$ ist.*

Andernfalls gäbe es ein f aus F, so daß $fp \not\subset U$; dann lägen aber auch fast alle fp_n außerhalb U, was unmöglich ist.

HILFSSATZ XIV. *Es sei a irgendein Punkt von K. Wenn die Folge*

$$a,\ a^2,\ a^3,\ \cdots,\ a^n,\ \cdots$$

(die „Potenzfolge von a") den Nullpunkt als Häufungspunkt besitzt, so konvergiert sie gegen diesen Punkt, und dasselbe gilt auch von der Potenzfolge eines jeden Punktes b, der zu einer passend gewählten Umgebung von a gehört.

Beweis. Für jede Umgebung G des Nullpunktes, deren abgeschlossene Hülle $\bar{G}$ kompakt ist, existiert vermöge Hilfssatz VIII ein solches m, daß

$$\bar{G}\, a^m \subset G$$

ist. Wegen Hilfssatz XII gilt die analoge Inklusion

$$\bar{G}\, b^m \subset G \tag{1}$$

für jeden Punkt b einer gewissen Umgebung $U(a)$ von a.

Wir wollen zeigen, daß unter der Bedingung (1) die Folge b^n gegen den Nullpunkt konvergiert. Damit wird der Hilfssatz XIV offenbar bewiesen sein.

Wiederholte Anwendung der Formel (1) gibt zunächst für jedes ganzzahlige q

$$\bar{G}\, b^{qm} \subset \bar{G}\, b^{(q-1)m} \subset \cdots \subset \bar{G}\, b^m \subset G. \tag{2}$$

Wenn $c \subset \bar{G}$, $c \neq 0$ ist, so gilt somit

$$c\, b^{qm} \subset G \subset \bar{G} \quad \text{und daher} \quad b^{qm} \subset c^{-1}\bar{G}.$$

Alle Punkte von der Form b^{qm} liegen somit in der kompakten Menge $c^{-1}\bar{G}$, so daß die Folge b^{qm} mindestens einen Häufungspunkt besitzt.

Wir nehmen jetzt an, die Folge besäße einen von Null verschiedenen Häufungspunkt. Dann gäbe es eine gegen diesen Punkt konvergierende Teilfolge b^{mq_k}, wobei man offenbar voraussetzen kann, daß $q_k - q_{k-1} = v_k$ mit k unendlich groß wird. Offenbar konvergiert dann die Folge $b^{v_k m}$ gegen den Einheitspunkt 1 von K. Andererseits haben wir nach (2)

$$\bar{G}\, b^{v_k m} \subset \bar{G}\, b^m \subset G.$$

Anwendung des Hilfssatzes XIII ergibt dann (indem man für das dortige U die abgeschlossene Menge $\bar{G}\, b^m$ wählt)

$$\bar{G} = \bar{G} \cdot 1 \subset \bar{G}\, b^m \subset G,$$

was unmöglich ist, denn K ist ja zusammenhängend.

Dasselbe gilt natürlich auch für jede Folge der Gestalt b^{mq+k} mit konstantem k. Da aber die ganze Folge b^n aus m Folgen der obigen Gestalt besteht (die man erhält, indem man k alle Werte von 0 bis $m-1$ gibt), konvergiert auch diese Folge gegen 0, w. z. b. w.

HILFSSATZ XV. *Wenn die Folge a^n eine divergente Teilfolge enthält, so ist auch die ganze Folge a^n divergent. Dasselbe gilt auch von der Potenzfolge eines jeden Elementes b aus einer gewissen Umgebung $U(a)$.*

Beweis. Die Folge a^{-n} hat nach Hilfssatz X den Nullpunkt unter ihren Häufungspunkten; folglich konvergiert nicht nur a^{-n}, sondern auch jede Folge b^n mit b aus einer gewissen Umgebung von a^{-1} gegen Null. Vermöge Hilfssatz V folgt hieraus, daß a^n bzw. b^{-n} divergiert.

HILFSSATZ XVI. *Der Körper K zerfällt in drei paarweise zueinander fremde Mengen λ, μ und ϱ, wobei ein Punkt zu λ, bzw. zu μ gehört, wenn seine Potenzfolge gegen Null konvergiert bzw. divergent ist. Die Menge ϱ besteht aus allen Punkten, deren Potenzfolgen keine divergenten Teilfolgen enthalten und trotzdem den Nullpunkt nicht zum Häufungspunkt haben.*

Beweis. Folgt aus Hilfssätzen XIV und XV.

HILFSSATZ XVII. *$\lambda+\varrho$ ist kompakt.*

Beweis. Man nehme an, die Punktfolge x_n aus $\lambda+\varrho$ sei divergent. Dann konvergiert x_n^{-1} wegen Hilfssatz X gegen Null und, da λ eine Umgebung des Nullpunktes ist, sind alle x_n^{-1} bis auf endlich viele in λ enthalten. Es sei $x_n^{-1} \subset \lambda$. Vermöge Hilfssatz V ist dann x_n^m divergent, so daß x_n zu μ gehört, was unserer Voraussetzung widerspricht.

HILFSSATZ XVIII. *Wenn $a \subset \lambda+\varrho$ ein mit jedem Element aus K vertauschbares Element ist, so gilt: $\overline{\lambda a} \subset \overline{\lambda}$.*

Beweis. Wenn $x \subset \lambda$ ist, so konvergiert die Folge $x, x^2, \cdots, x^n, \cdots$ gegen Null; da aber die Folge $a, a^2, \cdots, a^n, \cdots$ keine divergente Teilfolge enthält, so konvergiert auch die Folge $ax, a^2x^2, \cdots, a^nx^n, \cdots$ gegen Null. Da aber diese Folge wegen $ax = xa$ mit der Folge $xa, (xa)^2, \cdots, (xa)^n, \cdots$ zusammenfällt, ist auch $ax \subset \lambda$. Es ergibt sich also $\lambda a \subset \lambda$, wenn wir von beiden Seiten dieser Inklusion die abgeschlossene Hülle bilden, erhalten wir: $\overline{\lambda a} \subset \overline{\lambda}$.

HILFSSATZ XIX. *Sei q ein beliebiges, von Null verschiedenes Element von K und F eine kompakte Teilmenge von K. Dann existiert ein endliches System von Elementen $a_1, a_2, \cdots, a_m$ aus F, so daß es für jedes Element x aus F ein a_i gibt, so daß $x - a_i \subset \lambda q$ ist.*

Beweis. Im gegenteiligen Falle würde eine unendliche Folge

$$a_1, a_2, \cdots, a_n, \cdots \tag{1}$$

aus F existieren, so daß für $i \neq j$ stets $a_i - a_j \not\subset \lambda q$ ist. Dies ist aber unmöglich: die Folge (1) hat nämlich einen Häufungspunkt a und wenn

die Teilfolge a_{n_i} gegen a konvergiert, so konvergiert die Folge $a_{n_i} - a_{n_{i+1}}$ gegen Null, und daher ist für genügend großes i

$$a_{n_i} - a_{n_{i+1}} \subset \lambda q.$$

HILFSSATZ XX. *Wenn F eine kompakte Teilmenge von K und p ein beliebiges, von Null verschiedenes Element von K ist, dann gibt es für jede ganze Zahl k eine ganze Zahl N, so daß, wenn*

$$x_1, x_2, \cdots, x_N \tag{1}$$

beliebige N Elemente aus F sind, es unter ihnen k Elemente, etwa $x_{n_1}, x_{n_2}, \cdots, x_{n_k}$ gibt, so daß $x_{n_i} - x_{n_j} \subset \lambda p$ gilt für beliebiges i und j $(i, j \leqq k)$.

Beweis. Wegen Hilfssatz VIII bilden die λq bei beliebigem q ein vollständiges Umgebungssystem des Nullpunktes. Daher kann man q so bestimmen, daß aus $x_1 \subset \lambda q$, $x_2 \subset \lambda q$ folgt: $x_1 + x_2 \subset \lambda p$. Für das auf diese Weise bestimmte q bestimmen wir ein System von Elementen $a_1, a_2, \cdots, a_m$, das den Bedingungen des Satzes XIX genügt, und setzen $N = km$. Dann entspricht jedem Element x_i des Systems (1) ein Element a_r, so daß $x_i - a_r \subset \lambda q$ ist. Da aber $N = km$ ist, so gibt es ein solches Element a_r, dem auf diese Weise mindestens k Elemente des Systems (1) entsprechen, etwa $x_{n_1}, x_{n_2}, \cdots, x_{n_k}$. Auf diese Weise haben wir für $(i, j) \leqq k$: $x_{n_i} - a_r \subset \lambda q$, $a_r - x_{n_j} \subset \lambda q$ und daher $x_{n_i} - x_{n_j} = (x_{n_i} - a_r) + (a_r - x_{n_j}) \subset \lambda p$.

HILFSSATZ XXI. *Bezeichnen wir mit λ_n die Menge aller Elemente der Form $x_1 + x_2 + \cdots + x_n$, wobei $x_i \subset \lambda$ ist, dann ist:*

1. *λ_n eine offene Menge;*
2. *ihre abgeschlossene Hülle $\overline{\lambda}_n$ besteht aus allen Elementen der Form $x_1 + x_2 + \cdots + x_n$, wobei $x_i \subset \overline{\lambda}$ ist; $\overline{\lambda}_n$ ist daher kompakt;*
3. *$\lambda'_n = \overline{\lambda}_n - \lambda_n$ ist nicht leer.*

Beweis. 1. und 2. erkennt man unmittelbar, 3. folgt daraus, daß K zusammenhängend und nicht kompakt ist.

HILFSSATZ XXII. *Wenn $x \subset \lambda_r$ und $y \subset \lambda_s$ ist, so ist $x + y \subset \lambda_{r+s}$; wenn $x \subset \overline{\lambda}_r$ und $y \subset \overline{\lambda}_s$ ist, so ist $x + y \subset \overline{\lambda}_{r+s}$.*

HILFSSATZ XXIII. *Wenn $x \subset \lambda_r$ und $y \subset \overline{\lambda}_s$ ist, so ist $x + y \subset \lambda_{r+s}$.*

Beweis. Da $y \subset \overline{\lambda}_s$ ist, so gibt es für jede Umgebung G des Nullpunktes ein $z \subset G$, so daß $y - z \subset \lambda_s$ ist. Da aber λ_r eine offene Menge ist, so ist für genügend kleines G: $x + z \subset \lambda_r$, also

$$x + y = (x + z) + (y - z), \qquad x + z \subset \lambda_r, \qquad y - z \subset \lambda_s$$

und mithin nach Hilfssatz XXII: $x + y \subset \lambda_{r+s}$.

HILFSSATZ XXIV. *Für jedes ganzzahlige $k > 1$ gibt es ein $\omega_k \subset \overline{\lambda}$, so daß $k\omega_k \not\subset \overline{\lambda}_{k-1}$.*

Beweis. Sei $p \neq 0$ so bestimmt, daß aus $y_i \subset \lambda p$, $i = 1, 2, \cdots, k-1$ folgt: $y_1 + y_2 + \cdots + y_{k-1} \subset \lambda$. (Ein solches p existiert, da die λp ein vollständiges Umgebungssystem des Nullpunkts bilden.) Die ganze Zahl N sei entsprechend Hilfssatz XX bestimmt für p, k und $F = \bar{\lambda}$. Sei weiter $z \subset \lambda'_N$. (Infolge Hilfssatz XXI existiert ein solches z.) Dann ist

$$z = x_1 + x_2 + \cdots + x_N, \qquad x_n \subset \bar{\lambda}.$$

Auf Grund von Hilfssatz XX können wir annehmen, daß für die ersten k Elemente dieser Summe $x_i - x_j \subset \lambda p$ gilt $(i, j \leqq k)$. Wir erhalten:

$$z = (x_1 - x_k) + (x_2 - x_k) + \cdots + (x_{k-1} - x_k) + k x_k + x_{k+1} + x_{k+2} + \cdots + x_N.$$

Wegen $x_i - x_k \subset \lambda p$ $(i < k)$ ergibt sich:

$$x' = (x_1 - x_k) + (x_2 - x_k) + \cdots + (x_{k-1} - x_k) \subset \lambda.$$

Aus der Inklusion $x_n \subset \bar{\lambda}$ folgt weiter:

$$x'' = x_{k+1} + x_{k+2} + \cdots + x_N \subset \bar{\lambda}_{N-k}.$$

Folglich haben wir nach Hilfssatz XXIII:

$$x' + x'' \subset \lambda_{N-k+1}.$$

Hieraus und aus $k x_k \subset \bar{\lambda}_{k-1}$ ergäbe sich dann nach Hilfssatz XXIII: $z \subset \lambda_N$, was der Voraussetzung widerspricht. Wir erhalten also $k x_k \not\subset \bar{\lambda}_{k-1}$ und $\omega_k = x_k$.

HILFSSATZ XXV. *Der Körper K hat die Charakteristik Null.*

Beweis. Nehmen wir an, der Körper K habe eine Charakteristik $p \neq 0$. Dann gilt für beliebiges $x \subset K: px = 0$. Dies ist aber unmöglich, denn nach Hilfssatz XXIV gilt $p\omega_p \not\subset \bar{\lambda}_{p-1}$; $\bar{\lambda}_{p-1}$ enthält aber für $p > 1$ den Nullpunkt.

HILFSSATZ XXVI. *Wenn e die Einheit des Körpers K ist und m und n zwei ganze positive Zahlen sind $(m \geqq n)$, dann ist* $\bar{\lambda}_m \cdot \frac{e}{n} \supset \bar{\lambda}$.

Beweis. Jedes Element der Menge $\bar{\lambda} \cdot ne$ können wir in der Form $x + x + \cdots + x + 0 + 0 + \cdots + 0$ (x n-mal, 0 $(m-n)$-mal genommen), wobei $x \subset \bar{\lambda}$. Folglich ist nach Hilfssatz XXI $\bar{\lambda} \cdot ne \subset \bar{\lambda}_m$. Indem wir diese Inklusion mit $\frac{e}{n}$ multiplizieren, erhalten wir: $\bar{\lambda} \subset \bar{\lambda}_m \cdot \frac{e}{n}$.

HILFSSATZ XXVII. $\frac{m}{n} e$ *liegt in λ, ϱ oder μ, je nachdem ob $m < n$, $m = n$ oder $m > n$ ist.*

Beweis. Wir betrachten zunächst den Fall $m>n$. Wäre $\frac{m}{n}e \subset \lambda + \varrho$, so hätten wir nach Hilfssatz XVIII $\bar{\lambda}\cdot\frac{m}{n}e \subset \bar{\lambda}$, da $\frac{m}{n}e$ mit allen Elementen von K vertauschbar ist. Da $\omega_m \subset \bar{\lambda}$ ist, ergibt sich $\frac{m}{n}\omega_m \subset \bar{\lambda}$ und folglich $m\omega_m \subset \bar{\lambda}_n$; nach Hilfssatz XXIV wissen wir aber, daß $m\omega_m$ nicht in $\bar{\lambda}_{m-1}$ liegt. Daraus ergäbe sich aber $n>m-1$, also $n \geqq m$ (anderenfalls hätten wir nämlich $m\omega_m \subset \bar{\lambda}_n \subset \bar{\lambda}_{m-1}$), was der Voraussetzung widerspricht. Also ist $\frac{m}{n}e \subset \mu$.

Ist $m<n$, so ist $\frac{n}{m}e \subset \mu$ und folglich nach Hilfssatz X $\frac{m}{n}e \subset \lambda$.

Ist endlich $m=n$, so ist $\frac{m}{n}e = e \subset \varrho$.

HILFSSATZ XXVIII. *Die Folge*

$$e, 2e, \cdots, ne, \cdots \tag{1}$$

divergiert und die Folge $e, \frac{e}{2}, \cdots, \frac{e}{n}, \cdots$ *konvergiert gegen Null.*

Beweis. Anderenfalls gäbe es eine konvergente Teilfolge der Folge (1) $n_1 e, n_2 e, \cdots, n_k e, \cdots$ wobei die ganzen Zahlen n_k eine wachsende Folge bilden. Die Folge $(n_2-n_1)e, (n_3-n_2)e, \cdots, (n_k-n_{k-1})e, \cdots$ würde dann gegen Null konvergieren, was unmöglich ist, da alle Elemente $(n_k-n_{k-1})e$ nach Hilfssatz XXVII außerhalb λ liegen und λ eine Umgebung des Nullpunktes ist.

Aus der Divergenz der Folge $e, 2e, \cdots, ne, \cdots$ folgt nach Hilfssatz X, daß die Folge $e, \frac{e}{2}, \cdots, \frac{e}{n}, \cdots$ gegen Null konvergiert.

HILFSSATZ XXIX. *K enthält einen Körper R, der isomorph ist mit dem gewöhnlichen stetigen Körper der rationalen Zahlen.*

Beweis. Nach Hilfssatz XXV enthält K einen Körper R, der im rein algebraischen Sinn isomorph ist mit dem Körper der rationalen Zahlen; dabei schreibt sich jedes Element von R in der Form re, wobei r eine beliebige rationale Zahl ist. Nach Hilfssatz XXVIII bilden die offenen Mengen $\lambda\cdot\frac{e}{n}$ ein vollständiges Umgebungssystem des Nullpunktes im Körper K. Daraus folgt aber, daß die Durchschnitte der $\lambda\cdot\frac{e}{n}$ mit R ein vollständiges Umgebungssystem des Nullpunktes in R bilden. Diese Durchschnitte bestehen aber nach Hilfssatz XXVII aus denjenigen Elementen re, für die $|r|<\frac{1}{n}$ ist. R ist also tatsächlich isomorph mit dem stetigen Körper der rationalen Zahlen.

HILFSSATZ XXX. *Im Körper K gilt die Cauchysche Konvergenzbedingung, d. h. für die Konvergenz der Folge*

$$a_1, a_2, \cdots, a_n, \cdots$$

ist notwendig und hinreichend, daß es zu einer jeden Umgebung U des Nullpunktes eine ganze Zahl N gibt, so daß aus $r > N$, $s > N$ folgt $a_r - a_s \subset U$.

Beweis. Die Notwendigkeit der Bedingung erkennt man unmittelbar; beweisen wir, daß sie hinreichend ist. Wenn die Folge (1) dem Cauchyschen Konvergenzkriterium genügt, so gibt es ein N, so daß $a_{N+r} - a_N \subset \lambda$ ist; da $\bar{\lambda}$ kompakt ist, hat daher die Folge $a_{N+1} - a_N, a_{N+2} - a_N, \cdots, a_{N+k} - a_N, \cdots$ mindestens einen Häufungspunkt a und enthält keine divergente Teilfolge; infolgedessen hat die Folge (1) einen Häufungspunkt $a_N + a$; andererseits hat sie aber, wie man leicht erkennt, auch *höchstens* einen Häufungspunkt und enthält keine divergente Teilfolge. Sie ist also konvergent.

HILFSSATZ XXXI. *Der Körper K enthält einen Körper D, der isomorph ist mit dem gewöhnlichen stetigen Körper der reellen Zahlen; dabei ist jedes Element von D mit jedem Element von K vertauschbar.*

Beweis. Auf Grund der Hilfssätze XXIX und XXX überzeugt man sich leicht, daß die abgeschlossene Hülle $\bar{R} = D$ isomorph ist mit dem stetigen Körper der reellen Zahlen. Da jedes Element von R mit jedem Element von K vertauschbar ist, so gilt das gleiche auch für die abgeschlossene Hülle $D = \bar{R}$.

HILFSSATZ XXXII. *Sei A_n die Menge der Elemente aus K der Gestalt*

$$x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n,$$

wobei $x_1, x_2, \cdots, x_n$ gegebene Elemente von K sind und die variablen Koeffizienten α_i in D liegen. Wenn dann A_n nicht mit K zusammenfällt, so existiert ein Element $z \subset \varrho$, das nicht auf folgende Weise dargestellt werden kann:

$$z = x + z', \quad x \subset A_n, \quad z' \subset \lambda.$$

Beweis. Die Menge A_n ist abgeschlossen in K. Wenn daher $a \not\subset A_n$ ist, so gibt es unter den Umgebungen von a der Gestalt $a + \lambda\alpha$, $(\alpha \subset D)$, welche ein vollständiges Umgebungssystem von a bilden, eine solche, die zu A_n fremd ist. Setzen wir dann $a' = a\alpha^{-1}$, so ergibt sich, daß $a' + \lambda = (a + \lambda\alpha)\alpha^{-1}$ zu $A_n = A_n\alpha^{-1}$ fremd ist. Sei nun B_n die Menge der Elemente der Gestalt $x + z$, $x \subset A_n$, $z \subset \lambda$. Offenkundig ist B_n eine offene Menge und a' gehört nicht zu B_n. Da K zusammenhängend ist, existiert also mindestens ein Randpunkt von B_n. Sei b ein solcher Punkt und möge die Folge $x_k + z_k$ $(x_k \subset A_n, z_k \subset \lambda)$ gegen b konvergieren. Die z_k haben mindestens einen Häufungspunkt z und die x_k daher den Häufungs-

punkt $x = b - z$. Wegen $x \subset A_n$, $b = x + z$ kann z nicht zu λ gehören, als Häufungspunkt der Punkte $z_k \subset \lambda$ gehört es aber zu $\bar{\lambda}$; folglich ist $z \subset \varrho$. Wenn sich nun z in folgender Form darstellen ließe: $z = x' + z'$, $x' \subset A_n$, $z' \subset \lambda$, so ergäbe sich: $b = x + x' + z' = x'' + z'$, $x'' \subset A_n$. Dann müßte aber b zu B_n gehören, was unmöglich ist. Folglich genügt z allen Bedingungen des Hilfssatzes.

HILFSSATZ XXXIII. *Es existiert eine linear unabhängige endliche Basis*

$$x_1, x_2, \cdots, x_m \tag{1}$$

des Körpers K bezüglich D, d. h. jedes Element von K kann auf eine und nur eine Weise in der Form

$$\sum_{i=1}^{m} \alpha_i x_i \qquad (\alpha_i \subset D) \tag{2}$$

dargestellt werden.

Beweis. Sei $x_1 = e$. Nehmen wir an, wir hätten bereits in K ein System

$$x_1, x_2, \cdots, x_n \tag{3}$$

von bezüglich D linear unabhängigen Elementen derart, daß $x_i \subset \varrho$ und $x_i - x_j \not\subset \lambda$ für $i \neq j$. Entweder fällt die Menge aller Elemente der Form

$$\sum_{i=1}^{n} \alpha_i x_i \qquad (\alpha_i \subset D) \tag{4}$$

bereits mit K zusammen, oder man kann das System nach Hilfssatz XXXII zu einem System $x_1, x_2, \cdots, x_n, x_{n+1}$ erweitern, das denselben Bedingungen genügt. Auf Grund der Kompaktheit von ϱ und wegen Hilfssatz XX erkennt man, daß diese Erweiterung des Systems (3) nicht endlos fortgesetzt werden kann, und folglich gelangen wir schließlich zu einem System (1), welches eine linear unabhängige Basis von K bezüglich D bildet.

Der Hilfssatz XXXIII bedeutet, daß K isomorph ist mit einer Divisionsalgebra im Körper der reellen Zahlen im Sinne von Dickson. In diesem Falle muß K mit einem der drei in Satz I angegebenen Körper isomorph sein, was zuerst Frobenius bewiesen hat[6]. Ein Beweis dieses Satzes, der nicht den komplizierten Hilfsapparat von Frobenius benutzt, findet sich in dem Buche „Algebren und ihre Zahlentheorien" von Dickson (deutsche Ausgabe S. 46 Satz V).

[6] Journal für reine und angewandte Mathematik, Bd. 84.

ZUR BEGRÜNDUNG DER PROJEKTIVEN GEOMETRIE.[1]

VON A. KOLMOGOROFF IN MOSKAU.

Wir wollen hier auf eine interessante Anwendung des Pontrjaginschen Satzes über die stetigen algebraischen Körper[2] hinweisen. Bekanntlich erhält man, wenn man bei der Begründung der projektiven Geometrie nur die (dreidimensionalen) Verknüpfungsaxiome annimmt, eine große Menge von möglichen Raumtypen. Diese verschiedenen Typen von projektiven Räumen entsprechen den verschiedenen Typen der abstrakten algebraischen Körper; in der Tat, in jedem projektiven Raume kann man bekanntlich (auf Grund des Desargueschen Satzes) eine Streckenrechnung einführen, und dabei ist der so entstandene (im allgemeinen nichtkommutative) Streckenkörper bis auf einen Isomorphismus eindeutig durch den gegebenen Raum definiert; umgekehrt, wenn ein abstrakter algebraischer Körper gegeben ist, kann man einen und nur einen ihm entsprechenden Typus von projektiven Räumen konstruieren.

Um die gewöhnliche reelle projektive Geometrie zu charakterisieren, mußte man nach der klassischen Methode zwei neue Gruppen von Axiomen einführen, nämlich die Anordnungsaxiome und die Stetigkeitsaxiome. Erst danach konnte man durch Hinzufügung von komplexen Elementen die komplexe projektive Geometrie begründen. Jetzt aber gibt uns die abstrakte Topologie die Möglichkeit, Stetigkeitseigenschaften unabhängig von der Anordnung aufzufassen und folglich ohne die Anordnungsaxiome, welche nur für die reelle Geometrie gelten, die komplexe Geometrie in *direkter* Weise axiomatisch zu begründen. Im folgenden soll ein solcher direkter Weg angegeben werden.

Es seien drei Systeme von Elementen gegeben, welche als Punkte, Geraden und Ebenen bezeichnet werden und folgenden Bedingungen genügen:

I. Jedes von diesen drei Systemen bildet einen topologischen, zusammenhängenden, bikompakten Raum.

II. Die Punkte, Geraden und Ebenen genügen den gewöhnlichen projektiven Verknüpfungsaxiomen.

III. Die Verknüpfungsrelationen sind stetig: die durch zwei Punkte gehende Gerade ist eine stetige Funktion von diesen Punkten (wenn diese Punkte nicht zusammenfallen) usw.

---

[1] Received May 26, 1931.

[2] Siehe die vorstehende Arbeit von Pontrjagin.

Wenn man in einem I.—III. genügenden projektiven Raume auf einer Geraden drei beliebige (nicht zusammenfallende) Punkte als Null-, Eins- und Unendlichkeitspunkte wählt und die entsprechende Streckenrechnung definiert, so werden in dem so entstandenen Streckenkörper, wie man leicht beweist, alle Forderungen des Pontrjaginschen Satzes befriedigt. Folglich ist dieser Streckenkörper entweder dem Körper der reellen, oder dem Körper der komplexen Zahlen, oder dem Quaternionenkörper isomorph. Folglich existieren nur drei verschiedene Typen von Räumen, welche den Forderungen I.—III. genügen, nämlich der gewöhnliche reelle projektive Raum, der komplexe projektive Raum und der Raum, welchen man mit Hilfe der homogenen Koordinaten aus dem Quaternionenkörper konstruiert.

Die letzte Möglichkeit wird ausgeschlossen, wenn man den Pascalschen Satz (und folglich die Kommutativität der Multiplikation) als ein neues Axiom annimmt. Schließlich wird die komplexe Geometrie durch das Vollständigkeitsaxiom ausgezeichnet.

ON THE RESULTANT OF THREE DOUBLE BINARY FORMS.[1]

By Neal H. McCoy.[2]

The primary purpose of this note is to give a determinantal expression for the resultant of three double binary forms. By using certain covariants which vanish for the common solutions of the forms equated to zero, Moore[3] has obtained the resultant in the form of a determinant for a limited range of cases. We shall here give a short discussion of the resultant from the standpoint used by König[4] and then express the resultant in the general case as the quotient of a determinant by a power of a single coefficient of one of the forms.

1. **Definition and properties of the resultant.** Let us consider the three general double binary forms;

$$(1) \quad f_{m,\lambda}(x_1, x_2; y_1, y_2), \quad g_{n,\mu}(x_1, x_2; y_1, y_2), \quad h_{p,\nu}(x_1, x_2; y_1, y_2),$$

where by the notation $f_{m,\lambda}(x_1, x_2; y_1, y_2)$ we mean a polynomial of degree m in the variables x_1, x_2; of degree λ in the variables y_1, y_2; and homogeneous in each set of variables. We assume that each of the numbers $m, \lambda, n, \mu, p, \nu$, is greater than zero. We shall also denote the forms (1) simply by f, g and h respectively.

It will be convenient to have a notation for the non-homogeneous polynomials obtained from (1) by placing $x_2 = y_2 = 1$, $x_1 = x$, $y_1 = y$, in each form. Let us denote these non-homogeneous polynomials by f', g', h'.

In the forms (1), we assign the weight unity to x_2 and then give each coefficient of the forms a weight equal to the weight of the power product of which it is the coefficient. This weight we shall call *the weight of the coefficient with regard to* x. In like manner we assign the weight unity to y_2 and define the weight of each coefficient *with regard to* y.

We have now the following fundamental theorem.

Theorem 1. *There exists one and only one polynomial*[5] *in the coefficients of the general forms* (1), *which we denote by* R, *with the following properties:*

[1] Received January 12, and August 21, 1931. — Presented to the American Mathematical Society, Feb. 28, 1931.

[2] National Research Fellow.

[3] T. W. Moore, *Extended results in elimination*, these Annals, vol. 30 (1928) pp. 92–100.

[4] J. König, *Einleitung in die allgemeine Theorie der algebraischen Größen*, B. G. Teubner, Leipzig, 1903.

[5] By this we mean that if R_1 and R_2 are two polynomials with these properties, then $R_1 = c R_2$, where c is a numerical constant independent of the coefficients of the forms.

(a) $R \equiv 0 \pmod{f', g', h'}$.[6]

(b) *R is an irreducible function of the coefficients.*

(c) *R is homogeneous in the coefficients of each form separately, of degree $n\nu+p\mu$ in the coefficients of f, of degree $m\nu+p\lambda$ in the coefficients of g and of degree $m\mu+n\lambda$ in the coefficients of h.*

(d) *R is isobaric of weight $mn\nu+mp\mu+np\lambda$ with regard to x and isobaric of weight $m\nu\mu+n\nu\lambda+p\lambda\mu$ with regard to y.*

This polynomial, R, whose existence is asserted in this theorem is defined to be the resultant of the general forms (1). We thus define the resultant only to within a constant factor which for most purposes is of no consequence.

In a paper as yet unpublished the author has proved a theorem corresponding to Theorem 1 for the resultant of a system of forms homogeneous in any number of sets of variables, there being an arbitrary number of variables in each set.[7] We shall in this note merely outline the method of proof for the special case considered here. The theorem follows by some fairly simple modifications of the classical discussion of the resultant of a system of forms in one set of variables given by König.[8]

Let $S_1(y)$ be the ordinary resultant of f' and g' considered as polynomials in x; and $S_2(y)$ the resultant of f' and h' also considered as polynomials in x. Then by a known property of such resultants (König, p. 271), we have

$$S_1(y) \equiv 0 \pmod{f', g'},$$

and

$$S_2(y) \equiv 0 \pmod{f', h'}.$$

If now we let S denote the resultant of $S_1(y)$ and $S_2(y)$, we get

$$S \equiv 0 \pmod{S_1(y),\ S_2(y)},$$

that is

$$S \equiv 0 \pmod{f', g', h'}. \tag{2}$$

We now state a lemma whose proof follows as in König, pp. 267, 272.

---

[6] That is, R can be expressed in the form,

$$R \equiv \varphi_1 f' + \varphi_2 g' + \varphi_3 h',$$

where φ_1, φ_2, φ_3, are polynomials in x and y.

[7] For a statement without proof of the degrees and weights of such a resultant see Sylvester, *On the degree and weight of the resultant of a multipartite system of equations*, Proceedings of the Royal Society of London, vol. 12 (1862-63), pp. 674-676; Mathematical Papers, 2, pp. 329-330.

[8] Loc. cit.

LEMMA. *If there exists a polynomial, ψ, in the coefficients of f, g, h, such that*

$$\psi \equiv 0 \pmod{f', g', h'},$$

then there exists one and only one irreducible polynomial, R, with the same property and ψ is divisible by R. Also R is homogeneous in the coefficients of each form and isobaric with regard to both x and y.

This lemma together with relation (2) establishes the existence and uniqueness of the resultant and parts (a) and (b) of the theorem. There remains then only the parts (c) and (d) concerning the degrees and weights. It may now be shown by a detailed but fairly simple extension of the method used by König (pp. 276–284) that the resultant contains the term,

$$a^{n\nu+p\mu} R_1^m R_2^\lambda, \tag{3}$$

with at most a coefficient independent of the coefficients of the given forms. Here a is the coefficient of $x_1^m y_1^\lambda$ in f, R_1 is the ordinary resultant of $(g')_{x=0}$ and $(h')_{x=0}$, and R_2 is the resultant of $(g')_{y=0}$ and $(h')_{y=0}$. A calculation of the degrees and weights of this term gives the values stated in parts (c) and (d) of our theorem. There is no difficulty in the case of the degrees. Let us calculate, for example, the weight with regard to x. The coefficient a is of weight zero. Each term of $(g')_{x=0}$ is of weight n and each term of $(h')_{x=0}$ is of weight p with regard to x. Hence R_1 is of weight $n\nu+p\mu$ and R_1^m is of weight $m(n\nu+p\mu)$. Now R_2 is of the same weight with regard to x as weights are ordinarily assigned in the case of binary forms and hence R_2^λ is of weight $np\lambda$ with regard to x. Thus the term (3) is of weight $mn\nu+mp\mu+np\lambda$ with regard to x. A similar calculation shows that the weight with regard to y is also that required by the theorem. This completes the proof of the theorem.

2. **The resultant in determinant form.** Let us use the homogeneous forms (1) and suppose the notation so arranged that $\lambda \geqq \mu \geqq \nu$. We consider the following set of equations:

$$\begin{aligned}
x_1^{r_1} x_2^{n+p-r_1-1} y_1^{s_1} y_2^{\mu-s_1-1} f &= 0 \quad (r_1 = 0,1, \cdots, n+p-1)\ (s_1 = 0,1, \cdots, \mu-1),\\
x_1^{r_2} x_2^{m+p-r_2-1} y_1^{s_2} y_2^{\lambda-s_2-1} g &= 0 \quad (r_2 = 0,1, \cdots, m+p-1)\ (s_2 = 0,1, \cdots, \lambda-1),\\
x_1^{r_3} x_2^{m+n-r_3-1} y_1^{s_3} y_2^{\lambda+\mu-\nu-s_3-1} h &= 0 \quad (r_3 = 0,1, \cdots, m+n-1)\ (s_3 = 0,1, \cdots, \lambda+\mu-\nu-1).
\end{aligned} \tag{4}$$

We have here $\mu(n+p)$ equations arising from f, $\lambda(m+p)$ from g and $(m+n)(\lambda+\mu-\nu)$ from h. These equations involve a total of $(\lambda+\mu)$

$(m+n+p)$ unknowns which are the various power products of the variables of degree $m+n+p-1$ in x_1, x_2; and of degree $\lambda+\mu-1$ in y_1, y_2. There are thus $n(\mu-\nu)+m(\lambda-\nu)$ more equations than unknowns. We shall now prove the following theorem.

THEOREM 2. *Omit from the set of equations* (4) *the* $n(\mu-\nu)$ *equations of the first set where both* $r_1 \geqq p$ *and* $s_1 \geqq \nu$, *and the* $m(\lambda-\nu)$ *of the second where both* $r_2 \geqq p$ *and* $s_2 \geqq \nu$; *and denote by* D *the determinant of the coefficients of the unknowns in the remaining equations. If* c *denotes the coefficient of* $x_1^p y_1^\nu$ *in* h *and* R *is the resultant of the three general forms* (1), *then*

$$R = (1/c)^{n(\mu-\nu)+m(\lambda-\nu)} D. \tag{5}$$

We show first that D does not vanish for general forms. For this purpose consider the special forms,

$$\bar{f} = x_2^m y_2^\lambda, \qquad \bar{g} = x_1^n y_2^\mu + x_2^n y_1^\mu, \qquad \bar{h} = x_1^p y_1^\nu.$$

Each column of D is designated by a power product of degree $m+n+p-1$ in x_1, x_2; and of degree $\lambda+\mu-1$ in y_1, y_2. In each row of D arising from $\bar{h}$ we have one and only one element different from zero and it is unity. Hence we strike out the rows of D arising from $\bar{h}$ and the columns of D denoted by the power products which occur here with coefficients not zero, i. e. all those divisible by $x_1^p y_1^\nu$. The remaining determinant still has the value $\pm D$. Next we strike out the rows arising from $\bar{f}$ and the columns denoted by power products which here have coefficients different from zero. These are all the power products not divisible by $x_1^p y_1^\nu$ but divisible $x_2^m y_2^\lambda$.[9] Again our remaining determinant is $\pm D$. Finally we consider the rows of D arising from $\bar{g}$ together with the remaining columns which are those denoted by power products not divisible by either $x_1^p y_1^\nu$ or by $x_2^m y_2^\lambda$. A simple calculation shows that this set of power products may be written in the form

$$x_1^a x_2^{m+n+p-a-1} y_1^b y_2^{\lambda+\mu-b-1},$$

where either

$$a<p, \qquad b>\mu-1, \tag{6}$$

or

$$a>n+p-1, \qquad b<\nu. \tag{7}$$

Let ω denote this set of power products. The power products arising from $\bar{g}$ may be written in one of the forms,

[9] That each row of D arising from $\bar{f}$ has a non-zero element among this set of columns is clear from the fact that the multipliers of f divisible by $x_1^p y_1^\nu$ are excluded in the definition of D.

(8) $$x_1^{n+r_2}\, x_2^{m+p-r_2-1}\, y_1^{s_2}\, y_2^{\lambda+\mu-s_2-1}$$

or

(9) $$x_1^{r_2}\, x_2^{m+n+p-r_2-1}\, y_1^{\mu+s_2}\, y_2^{\lambda-s_2-1},$$

where either $r_2 < p$ or $s_2 < \nu$. For any value of r_2, s_2, satisfying this condition, either (8) or (9), but not both, is contained in the set ω. Also each one of the set ω may be put in either the form (8) or (9). Hence there is one and only one element different from zero in each row and column of our remaining determinant. Hence for these special forms, $D = \pm 1$ and therefore $D \neq 0$ for general forms.

It is readily seen that

(10) $$D \equiv 0 \pmod{f', g', h'}.$$

For let us place $x_2 = y_2 = 1$, $x_1 = x$, $y_1 = y$, in the equations (4). Then each column of D consists of elements which are coefficients of a given power product of x and y, of degree not exceeding $m+n+p-1$ in x and $\lambda+\mu-1$ in y. Suppose the last column of D consists of the constant terms in the equations (4). Multiply each column of D by the power product of which its elements are coefficients and add to the last column. Each element of the last column is now of one of the types $\varrho f'$, $\varrho g'$, $\varrho h'$, where ϱ is a power product of x and y. We have then the result (10) by expanding D in terms of the elements of the last column.

It follows from the lemma in the previous section that D contains R as a factor. Also D is clearly of degree $n\nu+p\mu$ in the coefficients of f, of degree $m\nu+p\lambda$ in the coefficients of g and of degree $(m+n)(\lambda+\mu-\nu)$ in the coefficients of h. From Theorem 1 (c), we have that

(11) $$D = TR,$$

where T is a polynomial of degree $n(\mu-\nu)+m(\lambda-\nu)$ in the coefficients of h alone. We proceed to a determination of T by means of the weights of D and R.

Each equation of (4), except those omitted, gives rise to a row of D. The weight, with regard to either x or y, of any coefficient in one of these equations is less than the weight of the power product of which it is the coefficient by an amount equal to the weight of the multiplier of the form which gives rise to the equation. Let us consider, for example, the equation,

$$x_1^{\alpha}\, x_2^{m+n-\alpha-1}\, y_1^{\beta}\, y_2^{\lambda+\mu-\nu-\beta-1}\, h = 0,$$

where α and β have definite values. The weights, with regard to x, of the coefficients of this equation are uniformly less than the weights of

the power products of which they are the coefficients by an amount $m+n-\alpha-1$.

Let A_x represent the sum of the weights with regard to x of all power products of the variables of degree $m+n+p-1$ in x_1, x_2; and of degree $\lambda+\mu-1$ in y_1, y_2. Also let L_x be the sum of the weights with regard to x of the multipliers of f in (4) (omitting those where $r_1 \geqq p$, $s_1 \geqq \nu$); M_x the sum of the weights with regard to x of the multipliers of g (omitting those where $r_2 \geqq p$, $s_2 \geqq \nu$); and N_x the sum of the weights with regard to x of the multipliers of h. Similarly by means of the subscript y we will denote these weights with regard to y.

Now each term in the expansion of D is a product of elements one and only one of which is taken from each row and each column. Hence D is isobaric of weight $A_x - L_x - M_x - N_x$ with regard to x and isobaric of weight $A_y - L_y - M_y - N_y$ with regard to y. We have

$$A_x = (\lambda+\mu)\sum_{s=0}^{m+n+p-1} s,$$

$$L_x = (\mu-\nu)\sum_{r_1=0}^{p-1}(n+p-r_1-1)+\nu\sum_{r_1=0}^{n+p-1}(n+p-r_1-1),$$

$$M_x = (\lambda-\nu)\sum_{r_2=0}^{p-1}(m+p-r_2-1)+\nu\sum_{r_2=0}^{m+p-1}(m+p-r_2-1),$$

$$N_x = (\lambda+\mu-\nu)\sum_{r_3=0}^{m+n-1}(m+n-r_3-1).$$

These are easily reduced to the forms,

$$A_x = \tfrac{1}{2}(\lambda+\mu)(m+n+p-1)(m+n+p),$$
$$L_x = \tfrac{1}{2}p(\mu-\nu)(2n+p-1)+\tfrac{1}{2}\nu(n+p)(n+p-1),$$
$$M_x = \tfrac{1}{2}p(\lambda-\nu)(2m+p-1)+\tfrac{1}{2}\nu(m+p)(m+p-1),$$
$$N_x = \tfrac{1}{2}(\lambda+\mu-\nu)(m+n-1)(m+n).$$

It follows that

$$A_x - L_x - M_x - N_x = mn\nu + mp\mu + np\lambda.$$

A similar calculation shows that

$$A_y - L_y - M_y - N_y = m\nu\mu + n\nu\lambda + p\lambda\mu.$$

We have then by Theorem 1 (d) that D is of the same weight as R and hence T in (11) must be of weight zero with regard to both x and y. But the only coefficient of h of weight zero with regard to both x and y is c, the coefficient of $x_1^p y_1^r$, and thus T must be a power of c. From

the consideration of the degrees above it follows that this power is $n(\mu-\nu)+m(\lambda-\nu)$, which completes the proof of the theorem.

It will be seen that in the special case $\lambda=\mu=\nu$, the expression (5) gives R in the form of a determinant without an extraneous factor. This is also clear from equations (4) as in this case we have the same number of equations as unknowns. Clearly the resultant is unchanged if we interchange x_1 and y_1, x_2 and y_2, in the given forms. Hence if $m=n=p$, we may also express the resultant as a determinant without extraneous factor.

We have given only one expression for the resultant and in this the coefficient of $x_1^p y_1^r$ in h plays an exceptional part. It is clear that by omitting a suitably chosen set of the equations (4) a corresponding expression for the resultant may be obtained in which the coefficient of any one of the terms, $x_1^p y_1^r$, $x_1^p y_2^r$, $x_2^p y_1^r$, $x_2^p y_2^r$, in h takes the place of c in an expression similar to (5).

PRINCETON UNIVERSITY.

ON THE COMPOSITION OF DEPENDENT ELEMENTARY ERRORS.[1]

By Cecil C. Craig.[2]

1. **Introduction.** In the mathematical theory of statistics one of the standard devices for the representation of a probability or a frequency function is to use its development in a series of Hermite polynomials under the supposition that the first few terms will give a satisfactory approximation. While the use of this series, called the type A series by Charlier, seems to be justified by the results in a wide class of cases, its justification on mathematical grounds has been very properly subjected to a great deal of criticism. The convergence of this and related series has been discussed by many investigators, but for the use mentioned above the convergence of the series is not of primary importance; rather the question is under what conditions a few terms may be expected to give a good approximation to the function in question and what improvement will be gained by adding one more term to those already used.

In a memoir of the greatest importance to the mathematical theory of statistics, Cramér[3] has studied the type A series and another related to it from this point of view and has given results of the kind that had hitherto been lacking. As the title of his paper indicates, he considers the probability variable x_n as the sum of n independent equal probability variables, $z_1, z_2, \cdots, z_n$, and he shows that, under certain quite general conditions on the probability functions of the component variables, the representations of the probability and frequency functions of x_n by means of either the type A series or a related series devised by him actually have asymptotic properties with respect to n.

It has long been known that under very general conditions the probability function of x_n approaches the normal probability function as n increases indefinitely. Certain Russian mathematicians, notably Markoff

[1] Received March 30, 1931.

[2] National Research Fellow.

[3] Cramér, H., On the Composition of Elementary Errors, First Paper; Skandinavisk Aktuarietidskrift (1928), pp. 13–74. Also see by the same author: Das Gesetz von Gauss und die Theorie des Risikos; ibid. (1923), pp. 209–237; On the Composition of Elementary Errors, Second Paper; ibid. (1928), pp. 141–180; On the Mathematical Theory of Risk, Försäkringsaktiebolaget Skandias Festskrift 1930 (Centraltryckeriet, Stockholm, 1930), pp. 7–84.

Unless otherwise noted all references to Cramér's work will be to the first paper listed above.

and Bernstein[4] have shown that the condition that the component variables be independent can to a very considerable extent be removed. Bernstein in the two papers cited above has given very important general results in this connection.

It is natural to ask to what extent the independence condition imposed by Cramér on the component variables can be removed without doing away with the asymptotic property of the type A series shown by him to exist. The present paper by a combination of the methods of Cramér and Bernstein shows that the independence condition can be replaced by conditions analogous to those imposed by Bernstein in the papers cited.

2. **Preliminary considerations and notation.** We shall in this paper in general follow the conventions and notation used by Cramér. Thus we shall call $V(t)$ the *probability function* of z if for every real t, $V(t)$ is equal to the probability that z is less than t plus half the probability that z has exactly the value t. Then $V(t)$ is a never decreasing function of t with $V(-\infty)=0$ and $V(\infty)=1$. For all values of t,

$$V(t) = \tfrac{1}{2}[V(t-0)+V(t+0)].$$

In case a finite derivative of $V(t)$ exists for all values of t, $V'(t)$ will be called the *frequency function* of t; the probability that z falls in the infinitesimal interval $(t, t+dt)$ is then $V'(t)\,dt$. Thus the *normal probability function* is defined by the equation,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,dt,$$

and the *normal frequency function* by,

$$\varphi(x) = \Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

The moments, "ordinary" and "absolute" of the probability function $V(t)$ are defined by

$$\alpha_\nu = \int_{-\infty}^{\infty} t^\nu\,dV(t), \quad \beta_\nu = \int_{-\infty}^{\infty} |t|^\nu\,dV(t), \quad \nu = 0, 1, 2, \cdots,$$

respectively, the right members being Stieltjes integrals. As a rule we shall consider probability functions for which it is assumed that only

[4] Bernstein, S., Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes, Math. Annalen, vol. 97 (1927), pp. 1–59; Sur les sommes de quantités dépendantes, Bulletin de l'Académie des Sciences de l'USSR (1926), pp. 1459–1478. See these two papers for references to Markoff's work.

Unless otherwise mentioned all references to Bernstein's work will be to the first paper listed.

a finite number of moments exist. The variance (the square of the standard deviation) of z is

$$\int_{-\infty}^{\infty} (t-\alpha_1)^2 \, dV(t) = \alpha_2 - \alpha_1^2.$$

Also the quantities ϱ_ν, defined by,

$$\varrho_\nu = \frac{\beta_\nu^{1/\nu}}{\beta_2^{1/2}},$$

play a very important part in the sequel.

We shall make considerable use of the semi-invariants of Thiele. For the semi-invariants γ_ν of the probability function $V(t)$ with the moments α_ν, we have the defining identity,

$$e^{\sum_1^\infty \frac{\gamma_\nu}{\nu!}\tau^\nu} = 1 + \sum_1^\infty \frac{\alpha_\nu}{\nu!}\tau^\nu,$$

with respect to τ which is to be regarded as purely formal.

For the probability function $W_n(x)$ of the sum $x_n = z_1 + z_2 + \cdots + z_n$, we shall use the special notations μ_ν, β_ν, and λ_ν for the moments and semi-invariants of x_n respectively. For the variance of $W_n(x)$ we have the special symbol σ^2. Thus

$$\sigma^2 = \mu_2 - \mu_1^2.$$

We also set,

$$F_n(x) = W_n(\sigma x),$$

in which $F_n(x)$ is a normalized probability function if we choose the origin for each z_i at the mean which entails no loss in generality, since then $F_n(x)$ satisfies the two conditions,

$$\int_{-\infty}^{\infty} x \, dF_n(x) = 0, \qquad \int_{-\infty}^{\infty} x^2 \, dF_n(x) = 1.$$

The adjuncts of the functions $V(x)$, $W_n(x)$, and $F_n(x)$ are complex functions which are defined for all real values of t by,

$$v(t) = \int_{-\infty}^{\infty} e^{-itx} \, dV(x), \qquad w(t) = \int_{-\infty}^{\infty} e^{-itx} \, dW_n(x),$$

$$f_n(t) = \int_{-\infty}^{\infty} e^{-itx} \, dF_n(x).$$

The type A series for $F_n(x)$ is formally,

$$F_n(x) = \Phi(x) + \frac{c_3}{3!}\Phi^{\mathrm{III}}(x) + \frac{c_4}{4!}\Phi^{\mathrm{IV}}(x) + \cdots,$$

in which,

$$(1) \qquad c_\nu = (-1)^\nu \int_{-\infty}^{\infty} H_\nu(x) \, dF_n(x),$$

where $H_\nu(x)$ is the ν-th Hermite polynomial defined by,

$$\frac{d^\nu}{dx^\nu} e^{-x^2/2} = (-1)^\nu H_\nu(x) e^{-x^2/2}.$$

(Since $F_n(x)$ is a normalized probability function, $c_1 = c_2 = 0$). If it be assumed that the frequency function $F_n'(x)$ exists, by formal differentiation we get,

$$F_n'(x) = \varphi(x) + \frac{c_3}{3!}\varphi^{\text{III}}(x) + \frac{c_4}{4!}\varphi^{\text{IV}}(x) + \cdots.$$

This is the development called by Charlier the type A series.

We shall denote by k the greatest index for which the moments α_ν and β_ν are assumed to be finite. With regard to k we shall, with Cramér, adopt the following conventions:

The letter θ will be used as a general symbol for a quantity, real or complex, which is always less in absolute value than a constant depending only on k.

The letter A will always denote a quantity which is less in absolute value than a constant depending only on k and the given function V.

For a fuller discussion of these and other preliminary considerations in the case of equal independent component variables see Cramér.[5]

In the present investigation we suppose that $x_n = z_1 + z_2 + \cdots + z_n$ is the sum of n dependent variables with the n-way probability function $V(z_1, z_2, \cdots, z_n)$. We assume for our developments that the component variables, $z_1, z_2, \cdots, z_n$ are *equal*, that is, that $V(z_1, z_2, \cdots, z_n)$ is symmetrical in the z_i's. It will be evident at the conclusion that the results obtained can be extended to the case of unequal component variables under conditions analogous to those indicated by Cramér for a similar extension in the case of independent components[6]. Then z_i taken apart from $z_1, z_2, \cdots, z_{i-1}, z_{i+1}, \cdots, z_n$ has a probability function which we shall denote by $V(t)$, with the moments α_ν, β_ν, the semi-invariants γ_ν, etc.: That is, $V(t)$ expresses the probability that z_i is less than t plus half the probability that z_i is exactly equal to t for the totality of sets of values of $z_1, z_2, \cdots, z_{i-1}, z_{i+1}, \cdots, z_n$ with which it may occur. In extension of the terminology used in the case of two dependent variables, $V(t)$ is called the *marginal* probability function of z_i and $\alpha_\nu, \beta_\nu, \gamma_\nu$, etc., the marginal moments, semi-invariants, etc.

Now let $V_i(z_i)$ be the probability function of z_i after values of $z_1, z_2, \cdots, z_{i-1}$ have been observed. Then $V_i(z_i)$ is a function of $z_1, z_2, \cdots, z_{i-1}$, as well as of z_i (though in a somewhat different sense), and so are its moments

[5] Loc. cit., pp. 13–38.

[6] Loc. cit., pp. 69–72.

and semi-invariants. These moments and semi-invariants, according to the sets of values of $z_1, z_2, \cdots, z_{i-1}$ which previously occur, fluctuate about the values α_ν, β_ν, and γ_ν. Let us suppose that this fluctuation of $\alpha_\nu^{(i)}$, defined by,

$$\alpha_\nu^{(i)} = \int_{-\infty}^{\infty} t^\nu \, d\, V_i(t),$$

never exceeds a_ν in absolute value where a_ν is independent of i. Also let the corresponding fluctuation of $\beta_\nu^{(i)}$ given by,

$$\beta_\nu^{(i)} = \int_{-\infty}^{\infty} |\, t\,|^\nu \, d\, V_i(t),$$

never exceed b_ν in absolute value where b_ν is also independent of i. Finally let c_ν be the greater of the two quantities, a_ν and b_ν.

3. **Concerning the expansion of $f_n(t)$ and $\log f_n(t)$.** Now we shall consider the expansion of the adjunct $w_n(t/\sigma) = f_n(t)$. We have,

$$f_n(t) = \int_{-\infty,\cdots,-\infty}^{\infty,\cdots,\infty} d\, V(z_1, z_2, \cdots, z_n)\, e^{-i\frac{t}{\sigma}\sum_1^n z_\nu},$$

the integral being an n-tuple Stieltjes integral. Then by the above

$$\begin{aligned} f_n(t) &= \int_{-\infty,\cdots,-\infty}^{\infty,\cdots,\infty} d\, V(z_1, z_2, \cdots, z_{n-1})\, e^{-i\frac{t}{\sigma}\sum_1^{n-1} z_\nu} \int_{-\infty}^{\infty} d\, V_n(z_n)\, e^{-i\frac{t}{\sigma} z_n} \\ &= \int_{-\infty,\cdots,-\infty}^{\infty,\cdots,\infty} d\, V(z_1, \cdots, z_{n-1})\, e^{-i\frac{t}{\sigma}\sum_1^{n-1} z_\nu} \\ &\quad \times \left[1 + \sum_1^{k-1} \frac{(\alpha_r + \vartheta_r^{(n-1)} c_r)}{r!}\left(-i\frac{t}{\sigma}\right)^r + \vartheta \frac{(\beta_k + \vartheta_k^{(n-1)} c_k)}{k!}\left(\frac{t}{\sigma}\right)^k\right] \end{aligned}$$

($|\vartheta_r^{(n-1)}| \leqq 1$, $r = 1, 2, \cdots, k$, and $|\vartheta| < 1$. In the sequel ϑ, with or without superscripts and subscripts, will be used as a general symbol for a quantity, real or complex, of absolute value $\leqq 1$.)

$$\begin{aligned} &= f_{n-1}(t)\left[1 + \sum_2^{k-1} \frac{\alpha_r}{r!}\left(-i\frac{t}{\sigma}\right)^r + \vartheta \frac{\beta_k}{k!}\left(\frac{t}{\sigma}\right)^k\right] \\ &\quad + \sum_2^{k-1} \frac{c_r}{r!}\left(-i\frac{t}{\sigma}\right)^r \int_{-\infty,\cdots,-\infty}^{\infty,\cdots,\infty} d\, V(z_1, \cdots, z_{n-1})\, \vartheta_r^{(n-1)}\, e^{-i\frac{t}{\sigma}\sum_1^{n-1} z_\nu} \\ &\quad + \frac{c_k}{k!}\left(\frac{t}{\sigma}\right)^k \int_{-\infty,\cdots,-\infty}^{\infty,\cdots,\infty} d\, V(z_1, \cdots, z_{n-1})\, \vartheta_k^{(n-1)}\, e^{-i\frac{t}{\sigma}\sum_1^{n-1} z_\nu}. \end{aligned}$$

(Since we have taken the origin for each z_i at the mean, the first order terms in the expansion of the adjunct $f_n(t)$ must vanish since the mean of a sum is the sum of the means.)

$$= f_{n-1}(t)\left[1+\sum_2^{k-1}\frac{\alpha_r}{r!}\left(-i\frac{t}{\sigma}\right)^r+\vartheta\frac{\beta_k}{k!}\left(\frac{t}{\sigma}\right)^k\right]$$
$$+\sum_2^{k-1}\vartheta_r^{(n-1)}\frac{c_r}{r!}\left(-i\frac{t}{\sigma}\right)^r+\vartheta_k^{(n-1)}\frac{c_k}{k!}\left(\frac{t}{\sigma}\right)^k.$$

Then by use of this recursion formula we get,

$$f_n(t) = \pi_n(t)+\pi_n(t)\sum_1^{n-1}\frac{\psi_\nu(t)}{\pi_{\nu+1}(t)}, \tag{2}$$

in which,

$$\pi_\nu(t) = \left[1+\sum_2^{k-1}\frac{\alpha_r}{r!}\left(-i\frac{t}{\sigma}\right)^r+\vartheta\frac{\beta_k}{k!}\left(\frac{t}{\sigma}\right)^k\right]^\nu = \left[v\left(\frac{t}{\sigma}\right)\right]^\nu,$$

and

$$\psi_\nu(t) = \sum_2^{k-1}\vartheta_r^{(\nu)}\frac{c_r}{r!}\left(-i\frac{t}{\sigma}\right)^r+\vartheta_k^{(\nu)}\frac{c_k}{k!}\left(\frac{t}{\sigma}\right)^k.$$

We know that,

$$\log\pi_n(t) = -n\frac{\beta_2}{2}\frac{t^2}{\sigma^2}+\sum_3^{k-1}\frac{n\gamma_r}{r!}\left(-i\frac{t}{\sigma}\right)^r+\theta n\beta_k\left(\frac{t}{\sigma}\right)^k.^{7} \tag{3}$$

Then since,

$$\log f_n(t) = \log\pi_n(t)+\log\left[1+\sum_1^{n-1}\frac{\psi_\nu(t)}{\pi_{\nu+1}(t)}\right], \tag{4}$$

we must consider the expansion of

$$\log\left[1+\sum_1^{n-1}\frac{\psi_\nu(t)}{\pi_{\nu+1}(t)}\right] = \log(1+\zeta),$$

and for this purpose we need $|\zeta|<1$.

Now for t confined to a finite interval $|\pi_\nu(t)|$ has a lower limit >0. In fact,

$$\left|v\left(\frac{t}{\sigma}\right)\right|^2 \geqq 1-\beta_2\frac{t^2}{\sigma^2},^{8}$$

and thus,

$$\log\left|v\left(\frac{t}{\sigma}\right)\right| \geqq \frac{1}{2}\log\left(1-\beta_2\frac{t^2}{\sigma^2}\right),$$

if $\beta_2\frac{t^2}{\sigma^2}<1$. For $0<x<1$,

$$\log(1-x) > -x-\frac{x^2}{2(1-x)},$$

[7] Cramér, H., loc. cit., p. 46, equation (45).

[8] Cf. Cramér, H., Das Gesetz von Gauss, etc., cited above, Art. 4.

which gives in our case,

$$\log\left|v\left(\frac{t}{\sigma}\right)\right| > \frac{-\beta_2 t^2}{2\sigma^2} - \frac{\beta_2^2 \frac{t^4}{\sigma^4}}{4\left(1-\beta_2\frac{t^2}{\sigma^2}\right)}.$$

Now let $\sigma = \sqrt{n}\,\pi_2\,\beta_2^{1/2}$; then,

$$\beta_2\frac{t^2}{\sigma^2} \leqq \left(\frac{\beta_3|t|^3}{\sigma^3}\right)^{2/3} = \left(\frac{\varrho_3^3|t|^3}{n^{3/2}\pi_2^3}\right)^{2/3},$$

since

$$\beta_2^{1/2} \leqq \beta_3^{1/3} \leqq \cdots \leqq \beta_k^{1/k}.^9$$

Thus if we take

$$|t| < \frac{1}{q}\,\frac{\sqrt{n}\,\pi_2}{\varrho_3},$$

with $q>1$, we have

$$\beta_2\frac{t^2}{\sigma^2} < \frac{1}{q^2} < 1.$$

(As we shall finally see, we can take $q=\sqrt{2}$). With this assumption on the range of t,

$$\log\left|v\left(\frac{t}{\sigma}\right)\right| > \frac{-\beta_2 t^2}{2\sigma^2} - \frac{q^2}{4(q^2-1)}\left(\frac{\beta_2 t^2}{\sigma^2}\right)^{1/2}\frac{\varrho_3^3|t|^3}{n^{3/2}\pi_2^3}$$

$$> \frac{-\beta_2 t^2}{2\sigma^2} - \frac{q}{4(q^2-1)}\,\frac{\varrho_3^3|t|^3}{n^{3/2}\pi_2^3},$$

and thus,

$$\left|v\left(\frac{t}{\sigma}\right)\right| > e^{-\frac{\beta_2 t^2}{2\sigma^2} - \frac{q}{4(q^2-1)}\frac{\varrho_3^3|t|^3}{n^{3/2}\pi_2^3}}.$$

This gives finally

$$|\pi_n(t)| > e^{-\frac{n\beta_2 t^2}{2\sigma^2} - \frac{q}{4(q^2-1)}\frac{\varrho_3^3|t|^3}{\sqrt{n}\,\pi_2^3}}, \tag{5}$$

for $|t| < \dfrac{1}{q}\,\dfrac{\sqrt{n}\,\pi_2}{\varrho_3}$. It is also to be observed that $|\pi_\nu(t)| \geqq |\pi_n(t)|$ for $\nu < n$.

Next, we can write

$$\sum_1^{n-1}\psi_\nu(t) = (n-1)\left[\sum_2^{k-1}\vartheta_r\frac{c_r}{r!}\left(-i\,\frac{t}{\sigma}\right)^r + \vartheta_k\frac{c_k}{k!}\left(\frac{t}{\sigma}\right)^k\right]. \tag{6}$$

---

[9] Cramér, H., loc. cit. Art. 15, Lemma 1.

Let us suppose that $\frac{c_r}{\beta_r} \leqq \omega^r$ *in which* $0 < \omega < 1$. Then

$$\left| \vartheta_r c_r \frac{t^r}{\sigma^r} \right| \leqq \left(\frac{\omega \beta_k^{1/k} |t|}{\sigma} \right)^r, \qquad k \geqq r.$$

Then for

$$|t| < \frac{1}{q} \frac{\sqrt{n}\, \pi_2}{\varrho_k} \leqq \frac{1}{q} \frac{\sqrt{n}\, \pi_2}{\varrho_2},$$

we have

$$\left(\frac{\omega \beta_k^{1/k} |t|}{\sigma} \right)^r < \frac{\omega^r}{q^r} < 1.$$

Thus we can write for t confined to this narrower interval

$$(7) \qquad \left| \sum_1^{n-1} \psi_r(t) \right| \leqq (n-1) \sum_2^{\infty} \frac{1}{r!} \left(\frac{\omega \beta_k^{1/k} |t|}{\sigma} \right)^r < (n-1)(e-2) \left(\frac{\omega \beta_k^{1/k} t}{\sigma} \right)^2.$$

Finally it will therefore be sufficient to have

$$\left| \frac{(n-1)\omega^2}{\pi_n(t)} \right| \leqq 1.$$

From (5) for $|t| < \frac{1}{q} \frac{\pi_2}{\varrho_k} \sqrt{n}$,

$$(8) \qquad |\pi_n(t)| > e^{-n\left[\frac{1}{2q^2\varrho_k^2} + \frac{1}{4q^2(q^2-1)}\right]} = e^{-n\chi_q}.$$

Then on setting $\omega < (n-1)^{-1/2} e^{-\frac{n}{2}\chi_q}$ we have

$$(9) \qquad |\zeta| < (e-2) \left(\frac{\beta_k^{1/k} t}{\sigma} \right)^2 < 0.8 < 1.$$

Thus we can expand $\log f_n(t)$ in powers of t under the conditions

1. $|t| < \frac{1}{q} \frac{\sqrt{n}\, \pi_2}{\varrho_k}$ with $q > 1$,
2. $\frac{c_r}{\beta_r} \leqq \omega^r$ in which $\omega < (n-1)^{-1/2} e^{-\frac{n}{2}\chi_q}$.

(As it will be seen below we shall finally take for ω the smaller value, $\omega_n = (n-1)^{-1/2} e^{-n\chi_q}$.)

The formal expansion of $f_n(t)$ in powers of $\frac{t}{\sigma}$ gives, as coefficients of the various powers, the moments of $F_n(x)$. The expansion of the first term on the right in (2) we know. The second member can be written,

$$\vartheta(n-1)\left[\sum_{2}^{k-1}\vartheta_r\frac{c_r}{r!}\left(-i\frac{t}{\sigma}\right)^r+\vartheta_k\frac{c_k}{k!}\left(\frac{t}{\sigma}\right)^k\right],$$

so that for our expansion we have,

$$f_n(t)=1+\sum_{2}^{k-1}\frac{[\mu_r+\vartheta_r(n-1)c_r]}{r!}\left(-i\frac{t}{\sigma}\right)^r+\frac{\vartheta B_k+\vartheta'(n-1)c_k}{k!}\left(\frac{t}{\sigma}\right)^k,$$

in which each ϑ_r is now real since in the exact expansion of $f_n(t)$ the coefficient of each power of it would be real. Therefore each "ordinary" moment of order r, $r=2, 3, \cdots, k-1$, of the sum $x_n=z_1+z_2+\cdots+z_n$ of our components with limited dependence differs by no more than $(n-1)c_r$ from the corresponding moment in case the components were independent.

If $\log f_n(t)$ is formally expanded in powers of $\frac{t}{\sigma}$ the coefficients give the semi-invariants of $F_n(x)$. Referring to (4) we see that we need the expansion of $\log(1+\zeta)$ in powers of $\frac{t}{\sigma}$. For small enough values of t we have already seen that this expression is convergent. And by the preceding paragraph in the same way that,

$$\log\pi_n(t)=\log\left[1+\frac{\mu_2}{2!}\left(-i\frac{t}{\sigma}\right)^2+\frac{\mu_3}{3!}\left(-i\frac{t}{\sigma}\right)^3+\cdots\right],$$

also

$$\log(1+\zeta)=\log\left[1+\vartheta_2(n-1)\frac{c_2}{2!}\left(-i\frac{t}{\sigma}\right)^2+\vartheta_3(n-1)\frac{c_3}{3!}\left(-i\frac{t}{\sigma}\right)^3+\cdots\right].$$

But since we have taken as hypothesis,

$$(n-1)c_r\leqq(n-1)\omega^r\beta_r<e^{-n\chi_0}\omega^{r-2}\beta_r<\omega^{r-2}\beta_r$$

we can write,

$$\log(1+\zeta)=\log\left[1+\frac{\vartheta_2\beta_2}{2!}\left(-i\frac{t}{\sigma}\right)^2+\vartheta_3\omega\frac{\beta_3}{3!}\left(-i\frac{t}{\sigma}\right)^3+\cdots\right]. \tag{10}$$

Then from the known expansion of $\log\left[1+\sum_{2}^{k}\frac{\alpha_r}{r!}\left(-i\frac{t}{\sigma}\right)^r\right]$,[10] in powers of $\frac{t}{\sigma}$, the sought expansion is readily written out. In fact, we have

[10] For the general term of this expansion see Frisch, R., Sur les semi-invariants et moments employés dans l'étude des distributions statistiques; Skrifter utgitt av det Norske Videnskaps-Akademi i Oslo, II. Hist.-Filos. Klasse (1926), No. 3, p. 20; or the author, An Application of Thiele's Semi-invariants to the Sampling Problem; Metron; Vol. VII, No. 4 (1928); p. 12.

(since in the case of independent components any semi-invariant of their sum is the sum of the corresponding semi-invariants of the components),[11]

$$\begin{aligned}
\lambda_2 &= n\gamma_2 + \vartheta_2\beta_2,\\
\lambda_3 &= n\gamma_3 + \vartheta_3\omega\beta_3,\\
\lambda_4 &= n\gamma_4 + \vartheta_4\omega^2\beta_4 - 3\vartheta_2^2\beta_2^2,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
\lambda_6 &= n\gamma_6 + \vartheta_6\omega^4\beta_6 - 15\vartheta_4\vartheta_2\omega^2\beta_4\beta_2 - 10\vartheta_3^2\omega^2\beta_3^2 + 30\vartheta_2^3\beta_2^3,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}$$

(up to and including λ_k).

Taking λ_6, for example, we have

$$\begin{aligned}
\lambda_6 &= (n\alpha_6 + \vartheta_6\omega^4\beta_6) - 15(n\alpha_4\alpha_2 + \vartheta_4\vartheta_2\omega^2\beta_4\beta_2) - 10(n\alpha_3^2 + \vartheta_3^2\omega^2\beta_3^2)\\
&\qquad + 30(n\alpha_2^3 + \vartheta_2^3\beta_2^3)\\
&= n\left[\left(\alpha_6 + \frac{\vartheta_6}{n}\omega^4\beta_6\right) - 15\left(\alpha_4\alpha_2 + \frac{\vartheta_4\vartheta_2}{n}\omega^2\beta_4\beta_2\right)\right.\\
&\qquad \left. - 10\left(\alpha_3^2 + \frac{\vartheta_3^2}{n}\omega^2\beta_3^2\right) + 30\left(\alpha_2^3 + \frac{\vartheta_2^3}{n}\beta_2^3\right)\right]\\
&= n\vartheta\left[\beta_6\left(1 + \frac{\vartheta_6}{n}\omega^4\right) + 15\beta_4\beta_2\left(1 + \frac{\vartheta_4\vartheta_2}{n}\omega^2\right)\right.\\
&\qquad \left. + 10\beta_3^2\left(1 + \frac{\vartheta_3^2}{n}\omega^2\right) + 30\beta_2^3\left(1 + \frac{\vartheta_2^3}{n}\right)\right]\\
&= \theta n\left(1 + \frac{\vartheta_6}{n}\right)\beta_6
\end{aligned}$$

and, in general, similarly,

$$\lambda_\nu = \theta n\left(1 + \frac{\vartheta_\nu}{n}\right)\beta_\nu, \qquad \nu = 2, 3, \cdots, k.$$

Let us set

$$1 + \frac{\vartheta_\nu}{n} = \pi_\nu^\nu, \qquad \nu = 2, 3, \cdots, k.$$

Then from the example of the calculation in the case of λ_6 it is seen that the θ in $\gamma_6 = \theta\beta_6$[12] is the same as the θ in $\lambda_6 = \theta n\pi_6^6\beta_6$. Thus we get, $\lambda_6 = n\pi_6^6\gamma_6$, and generally,

$$\lambda_\nu = n\pi_\nu^\nu\gamma_\nu, \qquad \nu = 2, 3, \cdots, k. \tag{11}$$

4. Further developments. After these preliminaries, it now only remains to be shown that the lines of development followed by Cramér

[11] Thiele, T. N., Theory of Observations, (C. & E. Layton, London, 1903), p. 39. Or see Cramér, H., loc. cit., Art. 15, Lemma 3.

[12] Cramér, H., loc. cit., Art. 15, Lemma 2.

in the case of independent components can also be followed in the case of components of limited dependence as above specified. It will not be necessary to go into details whenever they are the same as or quite analogous to those in the case of independent components; it will be sufficient to make clear what the developments of Cramér's articles 15–22 become in the present case. For that purpose it will perhaps be best to follow the same scheme of nine lemmas and three resulting theorems given there.

LEMMA 1. *We have* $\beta_1 \leqq \beta_2^{1/2} \leqq \beta_3^{1/3} \leqq \cdots \leqq \beta_k^{1/k}$.

It is only necessary to observe that this is true of the "absolute moments" of any probability function.

LEMMA 2. *For* $\nu = 2, 3, \cdots, k$, *is* $\gamma_\nu = \theta \beta_\nu$.

LEMMA 3. *For* $\nu = 2, 3, \cdots, k$, *is* $\lambda_\nu = n \pi_\nu^\nu \gamma_\nu$, *in which* $1 - \frac{1}{n} \leqq \pi_\nu^\nu \leqq 1 + \frac{1}{n}$.

This has just been shown in the section immediately preceding.

The symbolical polynomial $P_\nu(\Phi)$ is defined by the formal identity in t,

$$e^{\sum_1^{k-2} \frac{\lambda_{\nu+2}(-\Phi)^{\nu+2}}{\sigma^{\nu+2}(\nu+2)!} t^\nu} = 1 + \sum_2^{k-2} P_\nu(\Phi)\, t^\nu + N t^{k-1} + \cdots, \tag{12}$$

in which after expansion every power Φ^j is replaced by the derivative $\Phi^{(j)}$.[13]

LEMMA 4. *If we put according to* (12),

$$P_\nu(\Phi) = \sum_{j=1}^{\nu} H_{\nu j}\, \Phi^{(\nu+2j)}(x),$$

we have for $\nu = 1, 2, \cdots, k-2$,

$$H_{\nu j} = \theta \left(\frac{\varrho_k p_k}{\pi_2} \right)^{\nu+2j} n^{-\nu/2}$$

in which p_k *is defined as the least upper bound of* $1, \pi_2, \pi_3, \cdots, \pi_k$.

Using the modified form of Lemma 3 as given above, the proof is otherwise exactly the same as in the case of the corresponding lemma for independent components.

LEMMA 5. *For* $t \to 0$, *we have*

$$e^{t^2/2} f_n(t) = 1 + \sum_3^k \frac{c_\nu}{\nu!} (it)^\nu + o(t^k),$$

c_ν *being given by* (1).

The proof does not depend on the manner of composition of x.

The next lemmas which have to do further with the expansion of $f_n(t)$ require consideration in more detail.

[13] This is the definition of Cramér's polynomials P in which he also expands probability functions. See Cramér, loc. cit., Arts. 9–11.

LEMMA 6. *For* $|t| < \dfrac{\pi_2}{q\,\varrho_k\,p_k}\,n^{1/6}$, *in which* $\dfrac{1}{(1+1/n)^{1/2}} \leqq \dfrac{p_k}{\pi_2} \leqq \dfrac{(1+1/n)^{1/k}}{(1-1/n)^{1/2}}$,

$$(13)\quad e^{t^2/2} f_n(t) = 1 + \sum_3^{k-1} \frac{c_\nu}{\nu!}\left(-i\frac{t}{\sigma}\right)^\nu + \theta\, n^{-\varkappa/2}\left(\left|\frac{\varrho_k\,p_k\,t}{\pi_2}\right|^k + \left|\frac{\varrho_k\,p_k\,t}{\pi_2}\right|^{3(k-2)}\right)$$

$\left(\varkappa = \left[\dfrac{k+2}{3}\right]\right.$, *i. e., the largest integer in* $\left.\dfrac{k+2}{3}\right)$.

We had

$$(4)\qquad \log f_n(t) = \log \pi_n(t) + \log\left[1 + \sum_1^{n-1} \frac{\psi_\nu(t)}{\pi_{\nu+1}(t)}\right]$$

and it is known that

$$(14)\qquad \log \pi_n(t) = -n\frac{\beta_2\,t^2}{2\sigma^2} + \sum_3^{k-1} \frac{n\gamma_\nu}{\nu!}\left(-i\frac{t}{\sigma}\right)^\nu + \theta\, n\left(\frac{\varrho_k\,t}{\sqrt{n}\;\pi_2}\right)^k,$$ [14]

for $|t| < \dfrac{1}{\varrho_k}\,n^{1/6}$. And certainly $\dfrac{1}{q}\,\dfrac{\pi_2}{\varrho_k\,p_k}\,n^{1/6} < \dfrac{1}{\varrho_k}\,n^{1/6}$.

For the second term on the right in (4) we have seen that for $|t| < \dfrac{1}{q}\,\dfrac{\sqrt{n}\;\pi_2}{\varrho_k}$ which is $> \dfrac{1}{q}\,\dfrac{\pi_2}{\varrho_k\,p_k}\,n^{1/6}$ $(n>1)$ and for $\omega < (n-1)^{-1/2}\,e^{-n/2\chi_q}$,

$$\left|\sum_1^{n-1} \frac{\psi_\nu(t)}{\pi_{\nu+1}(t)}\right| = |\zeta| < 1.$$

Thus we can write

$$(15)\qquad \begin{aligned} \log(1+\zeta) &= \sum_{1 \leqq j < k/2} \frac{(-1)^{j+1}\,\zeta^j}{j} + \theta\,\zeta^{k/2} \\ &= \sum_{1 \leqq j < k/2} \frac{(-1)^{j+1}\,\zeta^j}{j} + \theta\,\frac{\beta_k\,t^k}{\sigma^k} \end{aligned}$$

by (9). From (7) we see that the series,

$$\sum_2^\infty \frac{1}{r!}\left(\frac{\beta_k^{1/k}|t|}{\sigma}\right)^r,$$

is a majorant for ζ and thus we have

$$\begin{aligned} \zeta^j &= \sum_{2j}^{k-1} \delta'_{\nu_j}\left(-i\frac{t}{\sigma}\right)^\nu + \theta \sum_k^\infty \frac{1}{\nu!}\left(\frac{j\,\beta_k^{1/k}|t|}{\sigma}\right)^\nu \\ &= \sum_{2j}^{k-1} \delta'_{\nu_j}\left(-i\frac{t}{\sigma}\right)^\nu + \theta\,\frac{\beta_k\,t^k}{\sigma^k}. \end{aligned}$$

[14] Cramér, H., loc. cit., p. 46, equation (45).

Setting this in (15) and combining the latter with (14), gives

$$\log f_n(t) = \sum_{2}^{k-1} \frac{\lambda_\nu}{\nu!}\left(-i\frac{t}{\sigma}\right)^\nu + (\theta n + \theta')\frac{\beta_k t^k}{\sigma^k}$$

or

$$\log\left[e^{t^2/2} f_n(t)\right] = \sum_{3}^{k-1} \frac{\lambda_\nu}{\nu!}\left(-i\frac{t}{\sigma}\right)^\nu + \theta n \pi_k'^k \left(\frac{\varrho_k t}{\sqrt{n}\,\pi_2}\right)^k, \tag{16}$$

in which $\pi_k'^k \leqq 1 + \frac{1}{n}$. (Now we take p_k to be the least upper bound of $1, \pi_2, \pi_3, \cdots, \pi_k, \pi_k'$).

Then for the values of t here considered,

$$\left|\sum_{3}^{k-1} \frac{\lambda_\nu}{\nu!}\left(-i\frac{t}{\sigma}\right)^\nu\right| < \theta n \sum_{3}^{\infty} \frac{1}{\nu!}\left(\frac{\varrho_k p_k |t|}{\sqrt{n}\,\pi_2}\right)^\nu$$

$$< \theta n \left(\frac{\varrho_k p_k |t|}{\sqrt{n}\,\pi_2}\right)^3 = \theta (\varrho_k p_k t)^2 \frac{\varrho_k p_k |t|}{\pi_2^3 \sqrt{n}} < \theta,$$

and

$$\left|\theta n \left(\frac{\varrho_k \pi_k' t}{\sqrt{n}\,\pi_2}\right)^k\right| < \left|\theta n \left(\frac{\varrho_k p_k t}{\sqrt{n}\,\pi_2}\right)^k\right| < \theta n^{1-k/3} < \theta.$$

Thus

$$e^{t^2/2} f_n(t) = e^{\sum_{1}^{k-3} \frac{\lambda_{\nu+2}(-it)^{\nu+2}}{\sigma^{\nu+2}(\nu+2)!}} + \theta n \left(\frac{\varrho_k p_k t}{\sqrt{n}\,\pi_2}\right)^k.$$

But as is apparent from the last few steps the remainder of the proof of this lemma is precisely similar to that in the case of independent components with the substitution of $\frac{\varrho_k p_k}{\pi_2}$ in place of ϱ_k.

Thus we also arrive, by the way, at

$$|c_\nu| < \theta \left(\frac{\varrho_k p_k}{\pi_2}\right)^\nu n^{-\nu/2+[\nu/3]} < M \left(\frac{\varrho_k}{\pi_2}\right)^\nu n^{-\nu/2+[\nu/3]} \tag{17}$$

in which M is independent of n.[15]

LEMMA 7. *For* $|t| < \frac{1}{4}\left(\frac{\pi_2}{\varrho_k p_k}\right)^3 \sqrt{n}$ *in which*

$$\frac{1}{(1+1/n)^{1/2}} \leqq \frac{p_k}{\pi_2} \leqq \frac{(1+1/n)^{1/k}}{(1-1/n)^{1/2}}$$

we have

$$|f_n(t)| < e^{-t^2/3}.$$

[15] Cf. Cramér, H., loc. cit., p. 49.

For dependent component variables

$$|f_n(t)|^2 = \int_{-\infty,\cdots,-\infty}^{\infty,\cdots,\infty} \int_{-\infty,\cdots,-\infty}^{\infty,\cdots,\infty} dV(y_1,\cdots,y_n)\, dV(z_1,\cdots,z_n) \times e^{-i\frac{t}{\sigma}\left(\sum_1^n y_i - \sum_1^n z_i\right)}$$

$$= \int_{-\infty,\cdots,-\infty}^{\infty,\cdots,\infty} \int_{-\infty,\cdots,-\infty}^{\infty,\cdots,\infty} dV(y_1,\cdots,y_{n-1})\, dV(z_1,\cdots,z_{n-1}) \times e^{-i\frac{t}{\sigma}\left(\sum_1^{n-1} y_i - \sum_1^{n-1} z_i\right)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dV_n(y_n)\, dV_n(z_n)\, e^{-i\frac{t}{\sigma}(y_n - z_n)}$$

$$\leqq \int_{-\infty,\cdots,-\infty}^{\infty,\cdots,\infty} \int_{-\infty,\cdots,-\infty}^{\infty,\cdots,\infty} dV(y_1,\cdots,y_{n-1})\, dV(z_1,\cdots,z_{n-1}) \times e^{-i\frac{t}{\sigma}\left(\sum_1^{n-1} y_i - \sum_1^{n-1} z_i\right)} \left[1 - (\beta_2 + \vartheta_2^{(n-1)} c_2)\frac{t^2}{\sigma^2} + \frac{4}{3}(\beta_3 + \vartheta_3^{(n-1)} c_3)\frac{|t|^3}{\sigma^3}\right] \tag{18}$$

$$= |f_{n-1}(t)|^2 \left[1 - \beta_2 \frac{t^2}{\sigma^2} + \frac{4}{3}\beta_3 \frac{|t|^3}{\sigma^3}\right] - \vartheta_2^{(n-1)} c_2 \frac{t^2}{\sigma^2} + \frac{4}{3}\vartheta_3^{(n-1)} \frac{|t|^3}{\sigma^3}.$$

Repetition gives,

$$|f_n(t)|^2 \leqq P_n(t)\left[1 + \sum_1^{n-1} \frac{U_\nu(t)}{P_{\nu+1}(t)}\right], \tag{19}$$

in which,

$$P_\nu(t) = \left[1 - \beta_2 \frac{t^2}{\sigma^2} + \frac{4}{3}\beta_3 \frac{|t|^3}{\sigma^3}\right]^\nu,$$

and,

$$U_\nu(t) = -\vartheta_2^{(\nu)} c_2 \frac{t^2}{\sigma^2} + \frac{4}{3}\vartheta_3^{(\nu)} c_3 \frac{|t|^3}{\sigma^3}.$$

Now

$$P_\nu(t) \geqq P_n(t), \qquad n > \nu,$$

and $P_n(t)$ has a lower bound >0 for t confined to a finite interval. In fact

$$P_n(t) > \left(1 - \beta_2 \frac{t^2}{\sigma^2}\right)^n$$

and by the same argument previously used,

$$P_n(t) > e^{-n\beta_2\frac{t^2}{\sigma^2} - \frac{q}{2(q^2-1)}\frac{\varrho_3^2|t|^3}{\sqrt{n}\pi_2}} = e^{-2n\gamma_3}, \qquad (q>1),$$

for $|t| < \dfrac{1}{q}\dfrac{\sqrt{n}\,\pi_2}{\varrho_3}$.

[16] Cf. Cramér, loc. cit.. Art. 17.

We can write

$$\sum_1^{n-1} U_\nu(t) = (n-1)\left[-\vartheta_2 c_2 \frac{t^2}{\sigma^2} + \frac{4}{3}\vartheta_3 c_3 \frac{t^3}{\sigma^3}\right].$$

Let us set,

$$1 + \sum_1^{n-1} \frac{U_\nu(t)}{P_{\nu+1}(t)} = \tau^2,$$

since from (18) it must be positive. Since for any $x > 0$,

$$\log x \leqq x - 1,$$

we get

$$2\log\tau \leqq \vartheta(n-1)\left[-\vartheta_2 c_2 \frac{t^2}{\sigma^2} + \frac{4}{3}\vartheta_3 c_3 \frac{t^3}{\sigma^3}\right] e^{2n\chi_q}$$
$$= -\vartheta_2\beta_2\frac{t^2}{\sigma^2} + \frac{4}{3}\vartheta_3\beta_3\frac{t^3}{\sigma^3},$$

if

$$\omega^2(n-1)e^{2n\chi_q} < 1.$$

Thus if we choose for ω a value $< \omega_n < (n-1)^{-1/2} e^{-n\chi_q}$ and if $|t| < \frac{1}{q}\frac{\sqrt{n}\,\pi_2}{\varrho_3}$, we have

$$\tau \leqq e^{-\vartheta_2\beta_2\frac{t^2}{2\sigma^2} + \frac{2}{3}\vartheta_3\beta_3\frac{t^3}{\sigma^3}}.$$

And since we know that,

$$|P_n(t)|^2 \leqq e^{-n\beta_2\frac{t^2}{\sigma^2} + \frac{4}{3}n\beta_3\frac{|t|^3}{\sigma^3}},$$ [17]

we have on multiplication,

$$|f_n(t)| \leqq e^{-\frac{t^2}{2} + \frac{2}{3}\left(1+\frac{\vartheta}{n}\right)n\beta_3\frac{|t|^3}{\sigma^3}} = e^{-\frac{t^2}{2} + \frac{2}{3}\frac{\varrho_3^3\pi_3''^3|t|^3}{\sqrt{n}\,\pi_2^3}}, \tag{20}$$

within this interval for t, in which the definition of π_3'' is obvious.

Finally taking $q \leqq 4$, we have concerning the range of t,

$$\frac{1}{4}\frac{\pi_2^3\sqrt{n}}{\varrho_k^3 p_k^3} \leqq \frac{\pi_2\sqrt{n}}{q\varrho_3},$$

if now p_k be taken as the least upper bound of $1, \pi_2, \pi_3, \pi_3'', \cdots, \pi_k, \pi_k'$. And for t within the narrower interval

$$1 - \frac{4}{3}\frac{\varrho_3^3\pi_3''^3|t|}{\sqrt{n}\,\pi_2^3} > \frac{2}{3}$$

which completes the proof of the lemma in question.

[17] Cramér, H., loc. cit., Art. 17.

If we define with Cramér,[18]

$$Q_n(x) = F_n(x) - \Phi(x) - \sum_3^{k-1} \frac{c_\nu}{\nu!} \Phi^{(\nu)}(x),$$

and

$$q_n(t) = f_n(t) - e^{-t^2/2} - \sum_3^{k-1} \frac{c_\nu}{\nu!} (it)^\nu e^{-t^2/2},$$

and introduce integrals of generalized order defined as follows: "The integral of order $\omega > 0$ of a function $g(x)$ will be defined by the equation

$$I^{(\omega)} g(x) = \frac{1}{\Gamma(\omega)} \int_{-\infty}^{x} (x-t)^{\omega-1} g(t)\, dt",$$

we have

LEMMA 8. *For* $0 < \omega \leqq k-1$, *we have*

$$I^{(\omega)} Q_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{itx}}{(it)^{\omega+1}} q_n(t)\, dt.$$

The proof of this lemma does not depend on the manner of composition of x.

LEMMA 9. *For* $0 < \omega \leqq k-1$, *we have*

$$(21)\quad I^{(\omega)} Q_n(x) = \frac{1}{\pi} R\left[\int_{\frac{1}{4}\left(\frac{\pi_2}{\varrho_k p_k}\right)^3 \sqrt{n}}^{\infty} \frac{e^{itx}}{(it)^{\omega+1}} f_n(t)\, dt\right] + \theta \left(\frac{\varrho_k p_k}{\pi_2}\right)^{3(k-2)} n^{-x/2}$$

$$(R[x] \equiv \text{real part of } x).$$

The terms containing the factor θ *have continuous first and second derivatives with respect to* x, *both of which are of the form* $\theta \left(\frac{\varrho_k p_k}{\pi_2}\right)^{3(k-2)} n^{-x/2}$.

By Lemma 8, we have,

$$I^{(\omega)} Q_n(x) = \frac{1}{\pi} R\left[\int_0^{\infty} \frac{e^{itx}}{(it)^{\omega+1}} q_n(t)\, dt\right] = A_1 + A_2 + A_3 + A_4,$$

in which,

$$A_1 = \frac{1}{\pi} R\left[\int_0^{\frac{1}{q}\frac{\pi_2}{\varrho_k p_k} n^{1/6}} \frac{e^{itx}}{(it)^{\omega+1}} q_n(t)\, dt\right],$$

$$A_2 = \frac{1}{\pi} R\left[\int_{\frac{1}{q}\frac{\pi_2}{\varrho_k p_k} n^{1/6}}^{\infty} \frac{e^{itx}}{(it)^{\omega+1}} \{q_n(t) - f_n(t)\}\, dt\right],$$

$$A_3 = \frac{1}{\pi} R\left[\int_{\frac{1}{q}\frac{\pi_2}{\varrho_k p_k} n^{1/6}}^{\frac{1}{4}\left(\frac{\pi_2}{\varrho_k p_k}\right)^3 n^{1/2}} \frac{e^{itx}}{(it)^{\omega+1}} f_n(t)\, dt\right],^{19}$$

[18] Cramér, H., loc. cit., Art. 18.

[19] $\frac{1}{q}\frac{\pi_2}{\varrho_k p_k} n^{1/6} < \frac{1}{4}\left(\frac{\pi_2}{\varrho_k p_k}\right)^3 n^{1/2}$ for all $n > 64\left(\frac{\varrho_k p_k}{\pi_2}\right)^6$, $(q > 1)$. Only a slight modification of the proof is required for smaller values of n.

and

$$A_4 = \frac{1}{\pi} R\left[\int_{\frac{1}{4}\left(\frac{\pi_2}{\varrho_k p_k}\right)^3 n^{1/2}}^{\infty} \frac{e^{itx}}{(it)^{\omega+1}} f_n(t)\, dt\right].$$

The remainder of the proof is just the same as in the case of independent components with $\frac{\varrho_k p_k}{\pi_2}$ everywhere substituted for π_2.

This completes the set of lemmas preliminary to the establishment of our theorems on the asymptotic expansions of probability and frequency functions in the case of limited dependence.

5. **Theorems on the expansion of probability and frequency functions.** First of all we give a theorem for the case of $k = 3$.

THEOREM I: *If $V(z_1, z_2, \cdots, z_n)$ is such that the marginal moment β_3 is finite, and $\frac{c_r}{\beta_r} \leqq \omega_n^r$, $r = 2, 3$, we have*

$$|F_n(x) - \Phi(x)| < 3\left(\frac{\varrho_3 p_3}{\pi_2}\right)^3 \frac{\log n}{\sqrt{n}},$$

in which $\frac{p_3}{\pi_2} \leqq \frac{(1+1/n)^{1/3}}{(1-1/n)^{1/2}}$, for $n>1$ and all values of x.

This result is the analog of Satz 1 of Cramér's paper, Das Gesetz von Gauss, etc., cited above. It is to be noted that in this paper the unit is different from that employed in his papers, On the Composition of Elementary Errors, and in the present one; to make comparisons one may regard the t of the "Das Gesetz" paper as $\sqrt{2}$ times the t of the later ones. And the ε^3 which appears in the first paper is defined by

$$\varepsilon^3 = \frac{n\beta_3}{n^{3/2}\beta_2^{3/2}} = \frac{\varrho_3^3}{\sqrt{n}},$$

for equal components in the present notation.

Inspection shows that to establish this analog of Satz 1, it is sufficient to show first, that

$$|f_n(t)| \leqq e^{-\frac{t^2}{2}+\frac{2}{3}\left(\frac{\varrho_3 p_3}{\pi_2}\right)^3 \frac{t^3}{\sqrt{n}}}$$

for $0<t<\frac{3}{8}\left(\frac{\pi_2}{\varrho_3 p_3}\right)^3 \sqrt{n}$.

But in (20) we have just this inequality for $0<t<\frac{1}{q}\frac{\pi_2}{\varrho_3 p_3}\sqrt{n}$ with $1<q \leqq 4$. If we choose further $q \leqq 8/3$, we have

$$\frac{3}{8}\left(\frac{\pi_2}{\varrho_3 p_3}\right)^3 \sqrt{n} \leqq \frac{3}{8}\frac{\pi_2}{\varrho_3 p_3}\sqrt{n},$$

and thus the first sufficient condition is satisfied.

The second sufficient condition is that

$$|f_n(t)| > e^{-\frac{t^2}{2} - \frac{1}{2\sqrt{2}}\left(\frac{\varrho_3 p_3}{\pi_2}\right)^3 \frac{t^3}{\sqrt{n}}}, \tag{22}$$

for $0 < t < \frac{\pi_2}{\sqrt{2}\,\varrho_3 p_3} n^{1/6}$.

In place of (17) we can write

$$|f_n(t)|^2 \geqq \int_{-\infty,\cdots,-\infty}^{\infty,\cdots,\infty} \int_{-\infty,\cdots,-\infty}^{\infty,\cdots,\infty} dV(y_1, \cdots, y_{n-1})\, dV(z_1, \cdots, z_{n-1})$$
$$\times e^{-i\frac{t}{\sigma}\left(\sum_1^{n-1} y_i - \sum_1^{n-1} z_i\right)} \left[1 - (\beta_2 + \vartheta_2^{(n-1)} c_2)\frac{t^2}{\sigma^2}\right]$$
$$= |f_{n-1}(t)|^2 \left[1 - \beta_2 \frac{t^2}{\sigma^2}\right] - \vartheta_2^{(n-1)} c_2 \frac{t^2}{\sigma^2}.$$

Then by repetition

$$|f_n(t)|^2 \geqq R_n(t)\left[1 + \sum_1^{n-1} \frac{T_\nu(t)}{R_{\nu+1}(t)}\right],$$

in which,

$$R_n(t) = \left(1 - \beta_2 \frac{t^2}{\sigma^2}\right)^n,$$

and

$$T_\nu(t) = -\vartheta_2^{(\nu)} c_2 \frac{t^2}{\sigma^2}, \qquad (\vartheta_2^{(\nu)} \text{ is real}).$$

We can write

$$\sum_1^{n-1} T_\nu(t) = -(n-1)\vartheta_2 c_2 \frac{t^2}{\sigma^2},$$

and since

$$R_n(t) = \left(1 - \beta_2 \frac{t^2}{\sigma^2}\right)^n,$$

we have for $|t| < \frac{1}{q}\frac{\pi_2}{\varrho_3}\sqrt{n} > \frac{1}{\sqrt{2}}\frac{\pi_2}{\varrho_2 p_3} n^{1/6}$, if q be taken finally $\leqq \sqrt{2}$, $(n > 1)$,

$$R_n(t) > e^{-2n\chi_q}.$$

Then also

$$\left|\sum_1^{n-1} \frac{T_\nu(t)}{R_{\nu+1}(t)}\right| < \left|-(n-1)\vartheta_2 \omega^2 \beta_2 \frac{t^2}{\sigma^2} e^{2n\chi_q}\right| < 1$$

for $\omega \leqq \omega_n < (n-1)^{-1/2} e^{-n\chi_q}$. Then for values of $\omega \leqq \omega_n$,

$$|f_n(t)|^2 > R_n(t)\left[1 - \vartheta_2 \beta_2 \frac{t^2}{\sigma^2}\right].$$

Thus

$$2\log|f_n(t)| > \log R_n(t) + \log\left(1 - \vartheta_2\beta_2\frac{t^2}{\sigma^2}\right).$$

But for $0 < t < \dfrac{n^{1/6}}{\sqrt{2}\,\varrho_3} > \dfrac{1}{\sqrt{2}}\dfrac{\pi_2}{\varrho_3 p_3}n^{1/6}$, we have

$$\frac{1}{2}\log R_n(t) > -\frac{n\beta_2 t^2}{2\sigma^2} - \frac{n\beta_3 t^3}{2\sqrt{2}\,\sigma^3}.$$ [20]

And similarly

$$\frac{1}{2}\log\left[1 - \vartheta_2\beta_2\frac{t^2}{\sigma^2}\right] > -\frac{\vartheta_2\beta_2 t^2}{2\sigma^2} - \frac{\vartheta_2^2\beta_3 t^3}{2\sqrt{2}\,\sigma^3}.$$

Whence on adding logarithms

$$|f_n(t)| > e^{-\frac{t^2}{2} - n\left(1+\frac{\vartheta_2^2}{n}\right)\frac{\beta_3 t^3}{2\sqrt{2}\sigma^3}},$$

and

$$|f_n(t)| > e^{-\frac{t^2}{2} - \frac{1}{2\sqrt{2}}\left(\frac{\varrho_3 p_3}{\pi_2}\right)^3\frac{t^3}{\sqrt{n}}},$$

if p_3 is now taken as the least upper bound of 1, π_2, π_3, π_3', π_3'', and π_3''' with $\pi_3'''^3 = 1 + \dfrac{\vartheta_2^2}{n}$ and if $0 < t < \dfrac{1}{\sqrt{2}}\dfrac{\pi_2}{\varrho_3 p_3}n^{1/6}$.

The third and last sufficient condition is that

$$|\operatorname{amp} f_n(t)| \leqq \frac{1}{4}\left(\frac{\varrho_3 p_3}{\pi_2}\right)^3\frac{t^3}{\sqrt{n}}, \tag{23}$$

for $0 < t < \dfrac{\pi_2}{\varrho_3 p_3}n^{1/6}$.[21]

Now from (6) and (10) with $k = 3$,

$$1 + \zeta = 1 - \vartheta_2\frac{\beta_2 t^2}{2\sigma^2} + \vartheta_3\omega\frac{\beta_3}{3!}\frac{t^3}{\sigma^3}.$$

Let

$$1 + \zeta = re^{i\varphi}.$$

Then

$$r\cos\varphi \geqq 1 - \vartheta_2\beta_2 t^2 - \beta_3\frac{|t|^3}{6\sigma^3},$$

---

[20] Cramér, H., Das Gesetz von Gauss, etc., p. 223.

[21] Referring to Cramér's, Das Gesetz von Gauss, etc., Art. 5, ... the third of the properties of $v(t)$ (in the notation used here) can be replaced by

$$|\operatorname{amp} v(t)| < \frac{1}{8\sqrt{2}}\varepsilon^3 t^3,$$

as may be seen from the calculation at the top of p. 225.

and for $|t| < \frac{1}{\sqrt{2}} \frac{\pi_2}{\varrho_3 p_3} n^{1/6}$,

$$\frac{\beta_2 t^2}{2\sigma^2} \leqq \frac{1}{2}\left(\frac{\beta_3 |t|^3}{\sigma^3}\right)^{2/3} < \frac{1}{4 n^{2/3} p_3^2} < \frac{1}{4}$$

for $n > 1$. Also similarly,

$$\frac{\beta_3 |t|^3}{6\sigma^3} < \frac{1}{12\sqrt{2}\, n\, p_3^3} < .06.$$

Thus

$$r \cos\varphi > 1 - .25 - .06 > \frac{2}{3}.$$

Also

$$|r \sin\varphi| \leqq \vartheta \frac{\beta_3}{6} \frac{t^3}{\sigma^3},$$

for $t > 0$. Whence

$$|\varphi| \leqq |\tan\varphi| < \frac{1}{4} \vartheta_3 \beta_3 \frac{t^3}{\sigma^3}.$$

We know that for $0 < t < \frac{1}{\sqrt{2}} \frac{n^{1/6}}{\varrho_3} \geqq \frac{1}{\sqrt{2}} \frac{\pi_2}{\varrho_3 p_3} n^{1/6}$

$$|\operatorname{amp} \pi_n(t)| < \frac{2}{9} n \beta_3 \frac{t^3}{\sigma^3} < \frac{1}{4} n \beta_3 \frac{t^3}{\sigma^3},$$[22]

and therefore

$$|\operatorname{amp} f_n(t)| < \frac{n}{4}\left(1 + \frac{\vartheta}{n}\right) \beta_3 \frac{t^3}{\sigma^3} \leqq \frac{1}{4}\left(\frac{\varrho_3 p_3}{\pi_2}\right)^3 \frac{t^3}{\sqrt{n}},$$

if now p_3 is finally taken as the least upper bound of 1, π_2, π_3, π_3', π_3'', π_3''' and π_3^{IV}, with $\pi_3^{IV^3} = 1 + \frac{\vartheta}{n}$ and if $0 < t < \frac{1}{\sqrt{2}} \frac{\pi_2}{\varrho_3 p_k} n^{1/6}$.

To see that these three conditions really are sufficient for the proof of Theorem I, write $\tau = \sqrt{2}\, t$ and $\bar{\varepsilon}^3 = \left(\frac{\varrho_3 p_3}{\pi_2}\right)^3 n^{-1/2}$; then the remainder of the proof is exactly analogous to that of Satz 1 of Das Gesetz von Gauss, etc. with the t and ε^3 which appear there everywhere replaced by τ and $\bar{\varepsilon}^3$.

Theorem I besides being of interest in its own right, is used in the proof of

THEOREM II. *If $V(z_1, z_2, \cdots, z_n)$ is such that the marginal moment β_k $(k > 3)$ is finite and $\frac{c_r}{\beta_r} \leqq \omega_n^r$, $r = 2, 3, \cdots, k$, we have for $n > 1$ and all values of x,*

$$(24)\quad I^{(x-3)}\left[F_n(x) - \Phi(x) - \sum_3^{k-1} \frac{c_\nu}{\nu!} \Phi^{(\nu)}(x)\right] = \theta\left(\frac{\varrho_k p_k}{\pi_2}\right)^{3(k-2)} (\log n)^{\frac{1}{2}} n^{-\varkappa/2},$$

[22] Cramér, H., Das Gesetz von Gauss, etc., p. 224.

in which

$$\frac{p_k}{\pi_2} \leqq \frac{(1+1/n)^{1/k}}{(1-1/n)^{1/2}} < e^{\frac{1}{kn}+\frac{1}{2n}+\frac{1}{4n(n-1)}}.$$

The proof is exactly analogous to that for the corresponding theorem in the case of independent components for $k>3$[23] with $\frac{\varrho_k p_k}{\pi_2}$ substituted everywhere for ϱ_k.

This gives only an asymptotic expansion of an integral of a certain order of the probability function $F_n(x)$. To get such an expansion of $F_n(x)$ itself, in the case of independent components, Cramér, introduced his condition ξ on $V(x)$, as follows.[24]

Since $V(x)$ is a never decreasing function, it can be uniquely represented as the sum of three never decreasing functions:[25]

$$V(x) = \xi(x)+\eta(x)+\zeta(x),$$

such that $\xi(-\infty)=\eta(-\infty)=\zeta(-\infty)=0$ and that,

1. $\xi(x) = \int_{-\infty}^{x} \xi'(t)\,dt = \int_{-\infty}^{x} V'(t)\,dt$ for all values of x.
2. $\eta(x)$, the "saltus function", is equal to the sum of the saltuses of $V(x)$ at all its points of discontinuity which have abscissas less than x. If x is the abscissa of a point of discontinuity, half the saltus at this point should be included in this sum.
3. $\zeta(x)$, the "singular function" is continuous and has "almost everywhere" a derivative equal to zero.

Then $V(x)$ is said to satisfy the condition ξ if the term $\xi(x)$ in the decomposition of $V(x)$ is not identically zero.

We shall also specify that $V_i(x)$, $i=1,2,\cdots,n$, shall satisfy the condition ξ and moreover shall satisfy it *uniformly*, as defined below.

We now have

$$V_i(x) = \xi_i(x)+\eta_i(x)+\zeta_i(x)$$

etc. And as before as $x\to\infty$, $\xi_i(x)$ tends to a limit which will be denoted by ξ_i^*. Then

$$0<\xi^*\leqq 1; \qquad \eta_i(\infty)+\zeta_i(\infty) = 1-\xi_i^*,$$

and

$$v_i(t) = \int_{-\infty}^{\infty} e^{-itx}\,d[\xi_i(x)+\eta_i(x)+\zeta_i(x)].$$

Further

$$|v_i(t)| \leqq \left|\int_{-\infty}^{\infty} e^{-itx}\,\xi_i'(x)\,dx\right| + 1-\xi_i^*.$$

[23] Cramér, H., loc. cit., Art. 20.

[24] Cf. Cramér, H., pp. 59, 60.

[25] See e. g., Lévy, Paul, Calcul des probabilités (Gauthier-Villars et Cie, Paris, 1925), pp. 142–149.

The first term on the right for any given $t>0$ is less than

$$\int_{-\infty}^{\infty} \xi_i'(x)\,dx = \xi_i^*$$

by a quantity we shall call δ_i. We say that the functions $V_i(x)$ satisfy the condition ξ uniformly if the δ_i's possess a lower bound $\delta>0$ independent of n.

Then for the expansion of the probability function we have:

THEOREM III. *If $V(z_1, z_2, \cdots, z_n)$ is such that $V_i(x)$, $i = 1, 2, \cdots, n$, satisfies the condition ξ uniformly, the marginal moment β_k is finite, and $\frac{c_r}{\beta_r} < \omega_n^r$, $r = 2, 3, \cdots, k$, we have*

$$F_n(x) = \Phi(x) + \sum_{3}^{k-1} \frac{c_\nu}{\nu!}\Phi^{(\nu)}(x) + A n^{-k/2} \tag{25}$$

for $k \geqq 3$, $n>1$ and all values of x.

From the above, it is clear that we can determine a number $b>0$ such that

$$\left|v_i\left(\frac{t}{\sigma}\right)\right| < e^{-b} \qquad i = 1, 2, \cdots, n,$$

for every $t \geqq \frac{1}{4}\left(\frac{\pi_2}{\varrho_k p_k}\right)^3 \sqrt{n}$. Then since,

$$f_n(t) = \int_{-\infty,\cdots,-\infty}^{\infty,\cdots,\infty} dV(z_1, \cdots, z_{n-1}) e^{-i\frac{t}{\sigma}\sum_1^{n-1} z_\nu} \int_{-\infty}^{\infty} dV_n(z_n)\, e^{-i\frac{t}{\sigma} z_n},$$

we see that we have

$$|f_n(t)| < e^{-bn} \tag{26}$$

for such values of t.

Thus by Lemma 9, noting that $\frac{p_k}{\pi_2}$ is bounded ($n>1$), we have

$$I^{(\omega)} Q_n(x) = A\left(\frac{e^{-bn}}{\omega n^{\omega/2}} + n^{-k/2}\right)$$

for $0<\omega<1$.

The remainder of the proof is just the same as it is for the corresponding theorem in the case of independent components.

If in the decomposition of $V_i(x)$, $\eta_i(x) \equiv \zeta_i(x) \equiv 0$, $i = 1, 2, \cdots, n$, we have for every x,

$$V_i(x) = \int_{-\infty}^{x} V_i'(t)\,dt. \tag{27}$$

Now for any given $t>0$, $|v_i(t)|$ differs from unity by a quantity δ_i and we shall say that the functions $V_i(x)$ satisfy (27) uniformly if the δ_i's possess a lower bound $\delta>0$ independent of n.

Then, finally, for the expansion of the frequency function, we have the following:

THEOREM IV. *If* $V(z_1, z_2, \cdots, z_n)$ *is such that* a) (27) *holds uniformly;* b) *the derivative* $V_i'(x)$, $i = 1, 2, \cdots, n$, *is uniformly of bounded variation in the whole domain* $-\infty < x < \infty$; c) *the marginal moment* β_k *is finite* $(k \geqq 3)$; *and* d) $\frac{c_r}{\beta_r} \leqq \omega_n^r$, $r = 2, 3, \cdots, k$, *then*

$$F_n'(x) = \varphi(x) + \sum_3^{k-1} \frac{c_\nu}{\nu!} \varphi^{(\nu)}(x) + A n^{-\varkappa/2} \tag{28}$$

for $n > 1$ *and all values of* x.

We have

$$\left| \int_{-\infty}^{\infty} d V_n(z_n)\, e^{-i\frac{t}{\sigma} z_n} \right| = \left| \frac{1}{i\, t/\sigma} \int_{-\infty}^{\infty} d V_n'(z_n)\, e^{-i\frac{t}{\sigma} z_n} \right| < \frac{\sigma}{|t|} \int_{-\infty}^{\infty} |d V_n'(z_n)|$$

$$\leqq \frac{M\sqrt{n}}{|t|},$$

in which M is finite and independent of n. This gives

$$|f_n(t)| < \left(\frac{M\sqrt{n}}{|t|} \right)^n.$$

The remainder of the proof is just as for the analogous theorem in the case of independent components.

STANFORD UNIVERSITY, CALIFORNIA.

ON THE REAL MULTIPLICATION OF ELLIPTIC FUNCTIONS.[1]

By E. T. Bell.

1. **The difference equation.** The *transcendental* solution of the problem of real multiplication of elliptic functions has been classic for over a century; the *algebraic* solution has not yet been achieved, although it has engaged the attention of many writers.[2]

The problem can be stated as follows. Let n be a constant integer greater than zero. Then (see, for example, Fricke, loc. cit., p. 202, second paragraph), the problem of real multiplication of elliptic functions is equivalent to that of calculating the denominator of $\operatorname{sn} nu$ as a polynomial in $\operatorname{sn}^2 u$, and the *algebraic* problem of real multiplication is equivalent to producing *explicit rational formulas for the coefficients of the several powers of* $\operatorname{sn}^2 u$ *in this polynomial.*

The integer n is fixed in all that follows, and hence it will not be carried as a superscript in the notation. Write $z \equiv \operatorname{sn}^2 u$, $\theta \equiv k^2$, where k^2 is the modulus of $\operatorname{sn} u$, and let $\mu(n) = (n^2-1)/2$ or $n^2/2$ according as n is odd or even. The denominator of $\operatorname{sn} nu$ is then of the form

$$1 + \sum_{\nu=1}^{\mu(n)} a_\nu(\theta)\, z^\nu$$

where $a_\nu(\theta) \equiv a_\nu$ is a polynomial of degree $\nu - 1$ at highest in θ,

[1] Received August 17, 1931.

[2] Abel's transcendental solution (1827) is in his Œuvres, vol. 1, pp. 279–291. To bring Kiepert's solution (J. f. Math., vol. 76, 1873, pp. 21–33) in terms of the Weierstrassian functions to the corresponding degree of completeness as that demanded here, it would be necessary to evaluate the determinant D_{n-1} and to give explicit formulas for the derivatives of the $\wp$-function as polynomials in $\wp, \wp'$. Of older papers on the problem, we cite 3 by Cayley, Collected Papers nos. 67, 93, 99, and the note, ibid. p. 589, where he says "the investigation seems to show that the integration (of the partial difference equation in the form used by him) cannot be effected in any tolerably simple form". See also Paper 45, especially pp. 299–300, and Paper 57. He remarks (p. 298) "it seems hopeless to continue this investigation any further". From a modern arithmetical point of view (modular systems), Kronecker discussed the algebraic structure of the polynomials in $\operatorname{sn}^2 z$, Werke, vol. 4, Papers nos. 29, 30 (1883), and Paper 31 (1883–1889), especially the last. A sufficient account of the elliptic function aspect of the problem is given by Fricke, Die elliptischen Funktionen und ihre Anwendungen, vol 2 (Teubner, 1922), pp. 196–209, or Briot et Bouquet, Fonctions elliptiques, (1875), pp. 516–539, or Königsberger, Elliptische Funktionen (1874), vol. 2, pp. 190–219.

$$a_\nu \equiv \sum_{j=0}^{\nu-1} \alpha(\nu, j)\, \theta^j,$$

in which the $\alpha(\nu, j)$ are *rational integers.* We write $\alpha(0, 0) = 1$ for the constant term in the denominator of $\operatorname{sn} nu$, and adopt the useful *convention that any sum* $\sum_a^b$ *in which* $b < a$ *is identically zero.* We shall express $\alpha(\nu, j)$ as *a polynomial in* n^2 *with rational coefficients.*

Jacobi's mixed differential-difference equation for a_ν is (Fricke, loc. cit., p. 208),

$$(2\nu+1)(2\nu+2)\, a_{\nu+1} + 4n^2\theta(1-\theta)\, \frac{da_\nu}{d\theta} + 4\nu[(n^2-\nu)\theta - \nu]\, a_\nu + (n^2-2\nu+1)(n^2-2\nu+2)\,\theta^2\, a_{\nu-1} = 0.$$

At this stage in calculating a_ν by recurrence, it is customary to reduce the equation by using the fact, which follows from elementary properties of elliptic functions, that the polynomial a_ν is reciprocal in θ. For our purpose it is necessary to avoid this simplification. Substituting directly for a_ν its expression stated above, and making some obvious reductions, we reach the following mixed difference equation for the $\alpha(\nu, j)$.

Let j, ν *be integers subject to the restrictions*

$$j > 0, \qquad \nu \geqq 0.$$

Define the functions F, H, A *by*

$$F(\nu, j) \equiv -\frac{2[jn^2 - (\nu+j-1)^2]}{(\nu+j)(2\nu+2j-1)},$$

$$H(j) \equiv -\frac{(n^2-2j+3)(n^2-2j+4)}{2j(2j-1)},$$

$$A(\nu, j-1) \equiv F(j, \nu)\,\alpha(\nu+j-1, j-1) + H(\nu+j)\,\alpha(\nu+j-2, j-1).$$

Then the $\alpha(\nu, j)$ *satisfy the equation*

$$\alpha(\nu+j, j) = A(\nu, j-1) + F(\nu, j)\,\alpha(\nu+j-1, j), \tag{1}$$

with the initial conditions

$$\begin{gathered} \alpha(x, y) = 0, \quad x<0; \qquad \alpha(x, y) = 0, \quad y<0; \\ \alpha(0, 0) = 1; \qquad \alpha(x, 0) = 0, \quad x>0; \qquad \alpha(1, 1) = 0; \\ \alpha(x, y) = 0, \quad y>x-1, \quad x>0; \\ \alpha(x, y) = \alpha(x, x-y), \quad x \geqq 0. \end{gathered} \tag{2}$$

The function $F(\nu, j)$ *above is defined for any integers* ν, j *such that* $\nu + j \neq 0$.

In this form the problem is completely solvable. At the first stage we replace $\alpha(\nu+j-1, j)$ in (1) by its value obtained from the left of (1) on changing ν into $\nu-1$, and repeat the process a sufficient number of times, until in the final right hand member the only $\alpha(x, j)$ occuring has the argument $x = j$. By (2), since $j>0$, this last term vanishes, and therefore, by the definitions of F, H, A, $\alpha(\nu+j, j)$ is expressed linearly in terms of functions $\alpha(x, j-1)$. The reduction of the linear expression gives the first stage of the solution, from which can be written down the value of $\alpha(1+\nu, 1)$. The second stage applies the first to each of the $\alpha(x, j-1)$, by changing j into $j-1$, and is repeated a sufficient number of times, until $\alpha(j+\nu, j)$ is wholly expressed linearly in terms of known functions $\alpha(y, 1)$.

2. **First stage of solution.** Proceeding as just outlined, we find

$$\alpha(j+\nu, j)$$
$$= A(\nu, j-1) + \sum_{s=1}^{\nu-1} F(\nu, j)F(\nu-1, j)\cdots F(\nu-s+1, j)A(\nu-s, j-1),$$

for all integers $j>0$, $\nu \geqq 0$. The product of F's is easily reduced from the definition of F, and we get finally, for the same j, ν,

$$\alpha(j+\nu, j) = \sum_{s=0}^{\nu-1} P(j, \nu, s)\cdot$$
$$\cdot[L(j, \nu, s)\alpha(j+\nu-s-1, j-1) + M(j, \nu, s)\alpha(j+\nu-s-2, j-1)], \tag{3}$$

in which P, L, M are defined by

$$P(j, \nu, 0) = 1,$$
$$P(j, \nu, s) \equiv \prod_{t=0}^{s-1} [jn^2 - (j+\nu-1-t)^2] \qquad (s>0);$$

$$L(j, \nu, s) \equiv \frac{(-1)^{s+1}2^{2s+1}(2j+2\nu-2s-2)!}{(j+\nu)(2j+2\nu-1)!}\{(\nu-s)n^2 - (j+\nu-s-1)^2\},$$

$$M(j, \nu, s) \equiv \frac{(-1)^{s-1}2^{2s-1}(2j+2\nu-2s-2)!}{(j+\nu)(2j+2\nu-1)!}\cdot$$
$$\cdot(n^2-2j-2\nu+2s+3)(n^2-2j-2\nu+2s+4), \quad (s \geqq 0).$$

3. **Value of $\alpha(1+\nu, 1)$.** In (3) take $j=1$, and refer to (2). Since $\alpha(0, 0) = 1$, $\alpha(x, 0) = 0$, $x \neq 0$, the first α on the right of (3) is not identically zero only when $s = \nu$, and the second only when $s = \nu-1$. Now $s = \nu$ is excluded by the limits of $\sum$; hence

$$\alpha(1+\nu, 1) = P(1, \nu, \nu-1)M(1, \nu, \nu-1) \qquad (\nu>0).$$

Referring to §§ 1, 2 for P, M, $\alpha(1,1)$, we get

$$\alpha(1,1) = 0,$$

$$\alpha(1+\nu,1) = \frac{(-1)^\nu 2^{2(\nu-1)}}{(1+\nu)(1+2\nu)!}\prod_{t=0}^{\nu}[n^2-(\nu-t)^2] \qquad (\nu>0). \tag{4}$$

4. **The operators** φ, ψ. The form of (3) suggests two bipartite operators by means of which the complete solution can easily be obtained. These operators are introduced chiefly to avoid complications in printing; if their equivalents, from the definitions given presently, be substituted in the final formulas, the explicit solution is thereby expressed fully in terms of elementary algebra. The operators are $\varphi \equiv (\delta, \sigma)$, $\psi \equiv (\delta, \tau)$, where δ, σ, τ are operators which we define first; each of φ^0, ψ^0, δ^0, σ^0, τ^0 denotes the identity operator.

Let p be an integer >0. Then φ^p denotes p applications of φ (or of φ^1), and similarly for the rest. In a product of operators, such as $\psi^{q_1}\varphi^{p_1}\cdots\psi^{q_i}\varphi^{p_i}$, the operators are applied in the order from right to left; thus φ^{p_i} is applied first, then ψ^{q_i} to the result, and so on. In all products of φ's and ψ's that occur, it will be noticed that neither two φ's nor two ψ's are adjacent in a product. The φ's, ψ's do not commute; the δ, σ, τ do.

Let p_i, q_i $(i = 1, \cdots)$ be integers >0, and x, y arbitrary integers. The s_{ij}, t_{ij} are positive integers, at first arbitrary, but later, from the notation in the formulas, summation blanks. The σ, s_{ij}, p_i, a_i occur together; likewise for the τ, t_{ij}, q_i, b_i; either the p_i or the q_i will occur with δ. These simple rules of notation are to avoid complicated summation limits.

We define

$$\delta^{p_i}y \equiv y - p_i,$$

$$\sigma^{p_i}x \equiv x - \sum_{a_i=1}^{p_i} s_{ia_i}, \qquad \tau^{q_i}x \equiv x - q_i - \sum_{b_i=1}^{q_i} t_{ib_i}.$$

Hence, by the preceding remarks, products of these operators are defined, since each of $\sigma^{p_i}x$, $\tau^{q_i}x$ is an integer, and hence each may be used as a new operand. For example,

$$\delta^{p_1}\cdots\delta^{p_i}y = y - \sum_{c=1}^{i} p_c;$$

$$\sigma^{p_1}\tau^{q_1}\cdots\sigma^{p_i}\tau^{q_i}x = x - \sum_{c=1}^{i}\left[\sum_{a_c=1}^{p_c} s_{ca_c} + q_c + \sum_{b_c=1}^{q_c} t_{cb_c}\right].$$

For *any* integer x we shall write the operator

$$\sum_{s_{i1}=0}^{\sigma^0 x-1}\sum_{s_{i2}=0}^{\sigma^1 x-1}\cdots\sum_{s_{ip_i}=0}^{\sigma^{p_i-1}x-1} \equiv \left\{\begin{matrix}\sigma^0 x-1, \cdots, \sigma^{p_i-1}x-1\\ s_{i1}, \cdots, s_{ip_i}\end{matrix}\right\}, \qquad \equiv \left\{\begin{matrix}\sigma^{p_i}x\\ s_i\end{matrix}\right\}.$$

Let $S(x, y, z)$, $T(x, y, z)$ be defined whenever x, y, z are integers. Then

$$\prod_{j=1}^{p_i} S(\delta^{j-1} y,\ \sigma^{j-1} x,\ s_{ij})$$

is defined. We now operate on this product with $\{\}$ as above, and write

$$\begin{Bmatrix} \sigma^{p_i} x \\ s_i \end{Bmatrix} \prod_{j=1}^{p_i} S(\delta^{j-1} y,\ \sigma^{j-1} x,\ s_{ij}) \equiv \begin{bmatrix} \sigma^{p_i} x \\ s_i \end{bmatrix}.$$

In precisely the same way, replacing (σ, s, p) by (τ, t, q) in the operator, and (S, s) by (T, t) in the operand, we define

$$\begin{Bmatrix} \tau^{q_i} x \\ t_i \end{Bmatrix} \prod_{j=1}^{q_i} T(\delta^{j-1} y,\ \tau^{j-1} x,\ t_{ij}) \equiv \begin{bmatrix} \tau^{q_i} x \\ t_i \end{bmatrix}.$$

If either of the symbols [] is followed by others in which the summation blanks s_{ij}, t_{ij} already used in the [] are implicit in the notation of the symbols adjoined, the summations in the [] extend over the entire expression. Such is the case, for example, if a symbol involving $\sigma^{p_i} x$ be adjoined to the first [].

Let $f(x, y)$ be defined whenever x, y are integers. The operators $\varphi \equiv (\delta, \sigma)$, $\psi \equiv (\delta, \tau)$ are now defined with respect to $f(y+x, y)$ by

$$\varphi^{p_i} f(y+x, y) \equiv \begin{bmatrix} \sigma^{p_i} x \\ s_i \end{bmatrix} f(\delta^{p_i} y + \sigma^{p_i} x,\ \delta^{p_i} y),$$

$$\psi^{q_i} f(y+x, y) \equiv \begin{bmatrix} \tau^{q_i} x \\ t_i \end{bmatrix} f(\delta^{q_i} y + \tau^{q_i} x,\ \delta^{q_i} y).$$

We shall need the results of any number of repetitions of these operations, in which either of φ, ψ is followed at a given step by the other. The result, $\psi^{q_i} \varphi^{p_i} f(y+x, y)$, of operating with ψ^{q_i} on $\varphi^{p_i} f(y+x, y)$ is, by definition, to pass the new operational factor ψ^{q_i} over the square bracket in $\varphi^{p_i} f(y+x, y)$ and operate with ψ^{q_i} on $f(\delta^{p_i} y + \sigma^{p_i} x,\ \delta^{p_i} y)$. The last can be written down from the second of the above definitions by replacing therein (x, y) by $(\sigma^{p_i} x,\ \delta^{p_i} y)$.

Now, referring to previous definitions, we see that the product $\tau^{q_i} \sigma^{p_i} x$ which occurs in the new square bracket symbol and the product $\tau^{q_i} \sigma^{p_i} x$ which enters as one of the arguments in the new f, play different parts; that for f is $\tau^{q_i} \sigma^{p_i} x$ as already defined for products of σ's and τ's; that in the bracket symbol states that the new bracket is formed by the same rule, operating on $\sigma^{p_i} x$, as that by which $\begin{bmatrix} \tau^{q_i} x \\ t_i \end{bmatrix}$ was formed by operating on x. To distinguish the bracket product we shall write $\tau^{q_i}(\sigma^{p_i} x)$, and

likewise in all similar cases: the *operand* is enclosed in (), and such enclosed operands are products of σ's and τ's having the meaning already defined (as in the argument of f).

We have thus,

$$\psi^{q_i}\varphi^{p_i}f(y+x,\,y) = \begin{bmatrix}\sigma^{p_i}x\\ s_i\end{bmatrix}\psi^{q_i}f(\delta^{p_i}y+\sigma^{p_i}x,\ \delta^{p_i}y),$$

$$= \begin{bmatrix}\sigma^{p_i}x\\ s_i\end{bmatrix}\begin{bmatrix}\tau^{q_i}(\sigma^{p_i}x)\\ t_i\end{bmatrix}f(\delta^{q_i}\delta^{p_i}y+\tau^{q_i}\sigma^{p_i}x,\ \delta^{q_i}\delta^{p_i}y).$$

At the next step we get

$$\varphi^{p_{i-1}}\psi^{q_i}\varphi^{p_i}f(y+x,\,y)$$

by passing $\varphi^{p_{i-1}}$ over both square brackets and operating on the f. The total resulting square bracket product is

$$\begin{bmatrix}\sigma^{p_i}x\\ s_i\end{bmatrix}\begin{bmatrix}\tau^{q_i}(\sigma^{p_i}x)\\ t_i\end{bmatrix}\begin{bmatrix}\sigma^{p_{i-1}}(\tau^{q_i}\sigma^{p_i}x)\\ s_{i-1}\end{bmatrix},$$

and the new f is

$$f(\delta^{p_{i-1}}\delta^{q_i}\delta^{p_i}y+\sigma^{p_{i-1}}\tau^{q_i}\sigma^{p_i}x,\ \delta^{p_{i-1}}\delta^{q_i}\delta^{p_i}y).$$

The next stage proceeds in the same way with $\psi^{q_{i-1}}$, and the law of formation is evident.

The formation of $\varphi^{p_i}\psi^{q_i}$, $\psi^{q_{i-1}}\varphi^{p_i}\psi^{q_i}$ is obvious from the above, and the results can be written down from the preceding by the substitution

$$(\sigma,\,\tau,\,s,\,t,\,p,\,q)\to(\tau,\,\sigma,\,t,\,s,\,p,\,q).$$

The binomial theorem for $\varphi+\psi$ must now be developed. If $P_1,\cdots,P_p$ are products of φ's and ψ's, we define $P_1+\cdots+P_p$ by

$$(P_1+\cdots+P_p)f(y+x,\,y) = P_1f(y+x,\,y)+\cdots+P_pf(y+x,\,y),$$

and $(\varphi+\psi)^pf(y+x,\,y)$, p an integer >0, is defined to be the result of operating precisely p times successively with $\varphi+\psi$ (or with $\varphi^1+\psi^1$) on $f(y+x,\,y)$. Thus we have

$$(\varphi+\psi)^1f(y+x,\,y) = (\varphi^1+\psi^1)f(y+x,\,y),$$
$$(\varphi+\psi)^2f(y+x,\,y) = (\varphi^2+\varphi^1\psi^1+\psi^1\varphi^1+\psi^2)f(y+x,\,y),$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$(\varphi+\psi)^pf(y+x,\,y) = \sum_{n=0}^{p}\binom{p}{n}\varphi^n\psi^{p-n}f(y+x,\,y),$$

in which $\binom{p}{n}\varphi^n\psi^{p-n}$ is the sum of the $\binom{p}{n}$ permutationally distinct products of precisely n operators φ and $p-n$ operators ψ formed from all the distinct permutations of these φ's and ψ's. For example,

$$\binom{4}{3}\varphi^3\psi^1 = \varphi\varphi\varphi\psi + \varphi\varphi\psi\varphi + \varphi\psi\varphi\varphi + \psi\varphi\varphi\varphi,$$
$$= \varphi^3\psi + \varphi^2\psi\varphi + \varphi\psi\varphi^2 + \psi\varphi^3.$$

An alternative definition of $\binom{p}{n}\varphi^n\psi^{p-n}$ in terms of the decompositions ($\equiv$ partitions, with attention to the order of the parts) of n is evident.

5. **Second stage of solution.** Returning to § 3 (3), and comparing with § 4, we set

$$(\nu, j) \equiv (x, y), \qquad \alpha(j+\nu, j) \equiv f(y+x, y),$$
$$P(j, \nu, s)\,L(j, \nu, s) \equiv S(y, x, s),$$
$$P(j, \nu, t)\,M(j, \nu, t) \equiv T(y, x, t).$$

Then (3) becomes

$$f(y+x, y) = (\varphi^1+\psi^1)f(y+x, y),$$

and if $y = 1$ the solution is complete. If $y>1$, we iterate precisely $y-1$ times. The resulting right hand member then expresses the left as a sum of products of known polynomials in n^2 (see (4), and the definitions of L. M.), and we have

$$f(y+x, y) = (\varphi+\psi)^{y-1}f(y+x, y) \qquad (y>1).$$

ELEMENTARY TRANSFORMATIONS.*

By Edgar R. Lorch.

I. Introduction.

In a series of remarkable memoirs, Liouville[1] demonstrated the impossibility of evaluating certain indefinite integrals, and of solving certain differential equations, in terms of elementary functions. The elementary functions are understood here to be those which are obtained in a finite number of steps by performing algebraic operations and taking exponentials and logarithms. Thus

$$z^2 \arcsin \sqrt{1-2\log_a z} + e^{(z\tan\sqrt[3]{1-z^3})^{1/2}}$$

in an elementary function of one variable.

A transformation on n variables is called an elementary transformation provided the n functions which define the transformation are each elementary. One of the problems arising in the study of elementary transformations is to determine under what conditions the inverse to such a transformation is itself elementary. For transformations operating on one variable, the problem has been completely solved by J. F. Ritt.[2]

In this paper we analyse this problem for the case of elementary transformations of a special type operating on two variables. In order to give meaning to the results which we have obtained, we recall some definitions. The exponential or logarithm of a non-constant algebraic function of u and v is a monomial of order one. A function which is algebraic in monomials of order one, and in u and v is a function of order one, provided it is not an algebraic function of u and v. An elementary transformation on two variables, such that the functions defining the transformation are either algebraic or of order one, at least one function being of order one, is called a transformation of order one. Our results deal with such transformations.

A transformation upon two variables defined by a pair of algebraic functions is called an algebraic transformation. If T is a transformation of order one, and A_1 and A_2 are any algebraic transformations, then $A_1 T A_2$ is a transformation of order one. We show in this paper that *if T is an elementary transformation of order one whose inverse is also elementary and*

* Received May 15, 1931.

[1] For a list of references, see J. F. Ritt, "Elementary Functions and their Inverses", Transactions of the American Mathematical Society, vol. 27 (1925), p. 68.

[2] Loc. cit.

of order one, there exist two algebraic transformations, A_1 and A_2, such that $A_1 T A_2$ is a transformation of one of the five following types:

(a) $$U = e^u, \qquad V = e^v;$$

(b) $$U = \log u, \qquad V = \log v;$$

(c) $$U = v^p e^u, \qquad V = qu + \log v.$$

Here p and q are rational and $pq \neq 1$.

(d) $$U = \theta, \qquad V = A(\theta, u, v);$$

$\theta = e^u$ or $\log u$; $A(\theta, u, v)$ is an algebraic function of θ, u, and v.

(e) $$U = A(\theta_1, \theta_2, \cdots, \theta_s; u, v), \qquad V = v;$$

$A(\theta_1, \cdots, \theta_s; u, v)$ is an algebraic function of $\theta_1, \cdots, \theta_s, u, v$. Each θ_i is either the exponential or the logarithm of some algebraic function of v only.

One may readily verify that the inverses of the above types of transformations are of order one. A transformation of order one may have an inverse which is elementary but not of order one. Thus the transformation

$$T: \quad U = e^u \log (u - v), \qquad V = e^v \log (u - v)$$

is of order one. Its inverse

$$T^{-1}: \quad U = \log u - \log \log \log \frac{u}{v}, \qquad V = \log v - \log \log \log \frac{u}{v}$$

is elementary but not of order one.

We shall not include in this paper a discussion of the function-theoretic considerations which underlie the work.[3] An acquaintance with the literature of elementary functions is not necessary to understand the exposition which follows except in a few instances which will be pointed out explicitly.

II. Some properties of elementary functions.

1. We recall briefly the classification of elementary functions. The variables u and v are monomials of order zero. Any algebraic function w of u and v defined by the equation

$$\alpha_0(u, v) w^n + \alpha_1(u, v) w^{n-1} + \cdots + \alpha_n(u, v) = 0, \tag{1}$$

where each $\alpha_i(u, v)$ represents a polynomial in u and v, is a function of order zero. The functions e^w and $\log w$, where w is a non-constant algebraic function, are monomials of ordre one. A function w satisfying an equation like (1) in which each $\alpha_i(u, v)$ is a polynomial in u, v, and monomials of order one is a function of order one, provided it is not an algebraic

[3] J. F. Ritt, loc. cit. Also ibid., vol. 25 (1923), p. 211.

function. The logarithm or exponential of a function of order one is a monomial of order two provided it is not a function of order zero or one. We proceed in this way, defining monomials and functions of all orders. Liouville, to whom we owe this classification, proved the existence of functions of all orders.

Given any function w of order n, it is apparent that the expression for it is by no means unique. But among all the expressions for w, there exist some which involve a number r of monomials of order n such that no expression of w involves fewer than r of these monomials. If the expression with which we are dealing contains exactly r monomials of order n, we say that w is expressed with a minimum number of monomials of order n. Let us designate the r monomials by $\theta_1^{(n)}, \theta_2^{(n)}, \cdots, \theta_r^{(n)}$. Since r is a minimum, it is clear that no one of these monomials can be expressed as an algebraic function of the remaining ones and of monomials of order less than n. A set of monomials of order n, $\theta_1^{(n)}, \theta_2^{(n)}, \cdots, \theta_s^{(n)}$, such that no $\theta_i^{(n)}$ is an algebraic function of the remaining $\theta_i^{(n)}$ and of monomials of order less than n will be defined to constitute an *independent set of monomials of order n* (or for short, an *independent set*). Any subset of an independent set is again an independent set.

2. Consider the partial derivative with respect to u of a function w of order n. Application of the ordinary rules for differentiation reveals that $\partial w/\partial u$ is an elementary function of order at most n, and that any monomial of order n which appears in the expression for $\partial w/\partial u$ appears also in that for w. In particular, if w is expressed with a minimum number of monomials of order n, the monomials of order n appearing in $\partial w/\partial u$ constitute an independent set. By induction, we can prove that the partial derivatives of w of all orders are elementary functions of order at most n, whose monomials of order n are found in the set of monomials of order n appearing in w.

3. Let $\theta_1^{(n)}, \theta_2^{(n)}, \cdots, \theta_r^{(n)}$ form an independent set of monomials of order n. Let us suppose that we have an algebraic function of the $\theta_i^{(n)}$ and monomials of lower orders which vanishes identically in u and v:

$$A(\theta_1^{(n)}, \theta_2^{(n)}, \cdots, \theta_r^{(n)}; \ \theta_1^{(n-1)}, \cdots; \ u, v) \equiv 0. \tag{2}$$

Then we may replace the $\theta_i^{(n)}$ in (2) by the independent variables z_i and the function

$$A(z_1, z_2, \cdots, z_r; \ \theta_1^{(n-1)}, \cdots; \ u, v) \tag{3}$$

will vanish dentically in the $r+2$ variables $z_1, \cdots, z_r, u, v$. For if the function (3) did not vanish identically, we could solve (2) for one of the monomials of order n, say, for $\theta_1^{(n)}$, thus showing that the monomials of order n appearing in (2) do not form an independent set. The rule of

procedure which states that in (2) we may replace the $\theta_i^{(n)}$ by the independent variables z_i without affecting the identity is known as Liouville's Principle. In the work which follows we make repeated application of it.

4. A function of the form

$$c_1 \log \varphi_1(u, v) + \cdots + c_m \log \varphi_m(u, v), \tag{4}$$

where each c_i is a constant, and $\varphi_i(u, v)$ is a function of order $n-1$ at most, will be called a *logarithmic sum of order* n, providing at least one $\varphi_i(u, v)$ is of order $n-1$.

Suppose that there exists in (4) a relation $\sum_{i=1}^{m} c_i p_i = 0$, the p_i being integers not all zero. Then we would have, for instance, $c_m = \sum_{i=1}^{m-1} c_i q_i$ with rational q_i. This enables us to express (4) as a logarithmic sum of order at most n, containing $m-1$ terms.

A set of numbers, $c_1, c_2, \cdots, c_m$, will be called *dependent or independent* according as there do or do not exist integers $p_1, p_2, \cdots, p_m$ not all zero, such that $\sum c_i p_i = 0$.

We now state four lemmas for elementary functions of two variables. The proofs, which we shall not give here, follow closely in each case those of the corresponding lemmas for the case of functions of one variable only.[4]

5. LEMMA. If in the expression for a function $\varphi(u, v)$ of order n, the number of exponentials of order n plus the number of logarithmic sums of order n is a minimum, each such exponential and logarithmic sum of order n is an algebraic function of $\varphi(u, v)$, a certain number of the partial derivatives of $\varphi(u, v)$, and of the monomials of order less than n which appear in the expression for $\varphi(u, v)$.

6. LEMMA. A function $\sum c_i \log \varphi_i(u, v)$ with no $\varphi_i(u, v)$ of order greater than $n-1$, with at least one $\log \varphi_i(u, v)$ of order n, and with independent c_i is a function of order n.

7. LEMMA. If $\varphi(u, v)$ and $\log \varphi(u, v)$ are both of order $n > 0$, $\varphi(u, v)$ is of the form $\xi_1(u, v) e^{\xi_2(u, v)}$ where $\xi_1(u, v)$ and $\xi_2(u, v)$ are each of order $n-1$.

LEMMA. If $\varphi(u, v)$ and $e^{\varphi(u, v)}$ are both of order $n > 0$, $\varphi(u, v)$ is of the form $\xi_1(u, v) + \log \xi_2(u, v)$ where $\xi_1(u, v)$ and $\xi_2(u, v)$ are each of order $n-1$.

8. We now prove the

LEMMA. *Let $\theta_1, \theta_2, \cdots, \theta_s$ constitute an independent set of monomials of order n. Let ζ, a monomial of order n, be an algebraic function of the θ_i and of monomials of order less than n. If ζ is a logarithm,*

[4] "Elementary Functions And Their Inverses", loc. cit.

(a) $$\zeta = q_1\theta_1 + q_2\theta_2 + \cdots + q_s\theta_s + \xi(u, v),$$

where each θ_i whose coefficient is not zero is a logarithm, q_i a rational number, and $\xi(u, v)$ a function of order $n-1$ at most. If ζ is an exponential,

(b) $$\zeta = \theta_1^{q_1} \cdot \theta_2^{q_2} \cdots \theta_s^{q_s} \cdot \xi(u, v),$$

where each θ_i whose exponent is not zero is an exponential, q_i a rational number, and $\xi(u, v)$ a function of order $n-1$ at most. Furthermore, for the special case $n = 1$, $\xi(u, v)$ in (a) *and* (b) *is a constant.*

The methods of Liouville may successfully be used to aid in the proof of this lemma, but we prefer to employ a method depending on the development of algebraic functions in infinite series which is more direct.[5]

We begin by proving (a). Let $\zeta = \log w$, w a function of order $n-1$. By hypothesis, $\log w$ is an algebraic function of the θ_i and of monomials of order less than n. We fix our attention upon some one θ_i which appears in the expression for $\log w$, say θ_1, and we write

(5) $$\log w = A(\theta_1, u, v).$$

This method of expression indicates that $\log w$ involves θ_1 algebraically and may involve other monomials. Let us assume that θ_1 is an exponential monomial, e^x. We differentiate (5) partially with respect to u, obtaining

(6) $$\frac{1}{w}\frac{\partial w}{\partial u} = A_{\theta_1}(\theta_1, u, v)\,\theta_1\frac{\partial x}{\partial u} + A_u(\theta_1, u, v),$$

in which the significance of the notation is obvious. Since the monomials of order n appearing in (6) constitute an independent set, we may, according to Liouville's Principle, replace θ_1 in (6) by an independent variable z without affecting the identity. Thus

(7) $$\frac{1}{w}\frac{\partial w}{\partial u} = A_z(z, u, v)\,z\frac{\partial x}{\partial u} + A_u(z, u, v).$$

Let us consider the function $A(z, u, v)$. We recall that in the neighborhood of $z = \infty$, we may develop this function into a converging infinite series of the following form:

(8) $$A(z, u, v) = \alpha_1 z^{q_1} + \alpha_2 z^{q_2} + \cdots.$$

The q_i are rational numbers with a common denominator such that $q_i < q_j$ for $i > j$. The α_i are algebraic functions of the monomials other than θ_1 which appear in the expression (5) for $\log w$.

[5] This method was introduced by J. F. Ritt. For an interesting application of it, see his note in the Comptes Rendus, t. 183, p. 331.

In order to find the expansion for $z = \infty$ of the function appearing in the right member of (7), we operate formally upon the series (8) in the manner indicated in (7). Hence, after rearranging terms, we have

$$\frac{1}{w}\frac{\partial w}{\partial u} = \left(q_1 \alpha_1 \frac{\partial x}{\partial u} + \frac{\partial \alpha_1}{\partial u}\right) z^{q_1} + \left(q_2 \alpha_2 \frac{\partial x}{\partial u} + \frac{\partial \alpha_2}{\partial u}\right) z^{q_2} + \cdots. \tag{9}$$

Here the coefficient of every power of z is identically zero in u and v. If $q_1 \neq 0$, we obtain from the coefficient of z^{q_1}

$$\theta_1 = e^x = e^{\delta} \cdot \alpha_1^{-1/q_1}, \tag{10}$$

where δ apparently involves v. We now differentiate (5) with respect to v. Repeating the steps which we have just outlined, we show quickly that δ is a constant. Since (10) expresses θ_1 as an algebraic function of $\theta_2, \cdots, \theta_s$ and of monomials of order less than n, we conclude that $q_1 = 0$. We then find that ζ is a linear function of α_1 with constant coefficients. Since α_1 does not involve θ_1, the expression for ζ may not involve any exponential of order n.

Let θ_1 be a logarithmic monomial, $\log x$. We differentiate (5) with respect to u, and in the result replace θ_1 by an independent variable z, obtaining

$$\frac{1}{w}\frac{\partial w}{\partial u} = A_z(z, u, v)\frac{1}{x}\frac{\partial x}{\partial u} + A_u(z, u, v). \tag{11}$$

From (8) we find that for large values of z, (11) may be expanded in the form

$$\frac{1}{w}\frac{\partial w}{\partial u} = \frac{\partial \alpha_1}{\partial u} z^{q_1} + q_1 \frac{\alpha_1}{x}\frac{\partial x}{\partial u} z^{q_1 - 1} + \cdots. \tag{12}$$

First, suppose $q_1 = 0$. From the coefficient of z^0 we find by integration that $\zeta = \alpha_1 + k$. We show quickly that k is a constant. If ζ actually involves θ_1, $q_1 \neq 0$, and α_1 is a constant. Suppose $q_1 \neq 1$. We set $q_j = q_1 - 1$ (possibly $\alpha_j \equiv 0$). Equating to zero the coefficient of $z^{q_1 - 1}$, we obtain by integration that θ_1 is a linear function of α_j. Since the θ_i form an independent set, we conclude that $q_1 = 1$. From the coefficient of z^0 in (12) we find that

$$\zeta = c_1 \theta_1 + \beta,$$

where c_1 is a constant and β is an algebraic function of $\theta_2, \cdots, \theta_s$ and monomials of order less than n. We repeat upon $\zeta - c_1 \theta_1$ the operations which we have just applied to ζ and in this way finally arrive at the equation

$$\zeta = c_1 \theta_1 + c_2 \theta_2 + \cdots + c_s \theta_s + \xi(u, v). \tag{13}$$

It is clear that $\xi(u, v)$ is of order at most $n-1$. In order to complete the proof of (a) in the lemma, we must show that each c_i is a rational number.

We assume henceforth that no c_i in (13) is zero. Let $\theta_i = \log x_i$. For the sake of convenience, suppose that $c_1, c_2, \cdots, c_r$ $(r \leqq s)$ are independent, while $c_j = \sum_{i=1}^{r} c_i q_{ij}$ with rational q_{ij} for $j = r+1, \cdots, s$. We shall prove that $r = 1$. We rewrite (13) in the form

$$\log w = c_1 \log x_1 \cdot x_{r+1}^{q_{1r+1}} \cdots x_s^{q_{1s}} + \cdots + c_r \log x_r \cdot x_{r+1}^{q_{rr+1}} \cdots x_s^{q_{rs}} + \xi(u, v).$$

If the numbers $1, c_1, c_2, \cdots, c_r$ were independent, we would have a contradiction of the lemma of § 6 since $\log w$ is of order n. Hence $1 = \sum_{i=1}^{r} c_i t_i$ with rational t_i. We may write

$$\xi(u, v) + \sum_{i=1}^{r} c_i \log \frac{x_i \cdot x_{r+1}^{q_{ir+1}} \cdots x_s^{q_{is}}}{w^{t_i}} = 0. \tag{14}$$

The functions whose logarithms are here taken are of order $n-1$ or less. The logarithms of these functions must be of order $n-1$ at most. Hence according to the first lemma of § 7, each such function may be expressed in the form $\psi_1(u, v) e^{\psi_2(u, v)}$, where $\psi_1(u, v)$ and $\psi_2(u, v)$ are of order less than $n-1$ (they may actually be constants). If we now assume that $r > 1$, we may eliminate w from two of the functions whose logarithms are taken in (14). We find that

$$x_1 = x_r^{p_r} \cdot x_{r+1}^{p_{r+1}} \cdots x_s^{p_s} \cdot \psi_1(u, v) \cdot e^{\psi_2(u, v)}, \tag{15}$$

where the p_i are rational, $\psi_1(u, v)$ and $\psi_2(u, v)$ are of order $n-2$ at most. But (15) indicates that θ_1 is an algebraic function of $\theta_r, \cdots, \theta_s$ along with monomials of order less than n. Hence $r = 1$. We have already seen that $1 = \sum_{i=1}^{r} c_i t_i$; this with $c_j = \sum_{i=1}^{r} c_i q_{ij}$ proves that the c_i are rational.

We now assume that ζ is an exponential, e^w, where w is of order $n-1$. We fix our attention upon some monomial, say θ_1, appearing in the expression for e^w and write (cf. (5))

$$w = \log A(\theta_1, u, v), \tag{16}$$

where the function $A(\theta_1, u, v)$ involves θ_1 algebraically; it possibly involves other monomials of the first n orders. Suppose first that θ_1 is a logarithmic monomial, $\log x$. We differentiate (16) with respect to u, and in the result replace θ_1 by z. We write the result in the form

$$\frac{\partial w}{\partial u} = \frac{\partial}{\partial z}[\log A(z, u, v)] \frac{1}{x} \frac{\partial x}{\partial u} + \frac{\partial}{\partial u}[\log A(z, u, v)]. \tag{17}$$

Let us consider the function $\log A(z, u, v)$. Comparison with (8) shows that for large values of z

$$(18)\quad \log A(z, u, v) = \log \alpha_1 + q_1 \log z + \log\left(1 + \frac{\alpha_2}{\alpha_1} z^{q_2 - q_1} + \frac{\alpha_3}{\alpha_1} z^{q_3 - q_1} + \cdots\right).$$

Having fixed a point (u, v), we may choose z sufficiently large so that

$$\left|\frac{\alpha_2}{\alpha_1} z^{q_2 - q_1} + \frac{\alpha_3}{\alpha_1} z^{q_3 - q_1} + \cdots\right| < 1.$$

For such values of z, we may develop the last term of (18) according to the formula for the expansion of $\log(1+x)$. We find that

$$(19)\quad \log A(z, u, v) = \log \alpha_1 + q_1 \log z + \beta_1 z^{r_1} + \beta_2 z^{r_2} + \cdots,$$

where the β_i are rational functions of the α_i and the r_i negative fractions with a common denominator such that $r_i < r_j$ for $i > j$.

In order to obtain the expansion of (17) around $z = \infty$, we operate upon (19) in the manner prescribed in (17). After rearranging the terms, we have

$$(20)\quad \frac{\partial w}{\partial u} = \frac{1}{\alpha_1}\frac{\partial \alpha_1}{\partial u} + \frac{q_1}{x}\frac{\partial x}{\partial u} z^{-1} + \frac{\partial \beta_1}{\partial u} z^{r_1} + r_1 \frac{\beta_1}{x}\frac{\partial x}{\partial u} z^{r_1 - 1} + \cdots.$$

The coefficient of each power of z in (20) must vanish identically in u and v. From the coefficient of z^0 we obtain

$$e^w = \alpha_1 \cdot e^{\delta}.$$

Here δ apparently involves v, but we find quickly that it is a constant. Since α_1 does not involve θ_1, we see that the expression for ζ does not involve any logarithms of order n.

Suppose now that θ_1 is an exponential monomial, e^x. We find that for z arbitrary

$$\frac{\partial w}{\partial u} = \frac{\partial}{\partial z}[\log A(z, u, v)]\, z \frac{\partial x}{\partial u} + \frac{\partial}{\partial u}[\log A(z, u, v)].$$

Comparing this with (19), we have

$$\frac{\partial w}{\partial u} = \frac{1}{\alpha_1}\frac{\partial \alpha_1}{\partial u} + q_1 \frac{\partial x}{\partial u} + \left(r_1 \beta_1 \frac{\partial x}{\partial u} + \frac{\partial \beta_1}{\partial u}\right) z^{r_1} + \cdots.$$

Since each r_i is negative, we obtain from the coefficient of z^0

$$\zeta = k \alpha_1 \theta_1^{q_1}.$$

It is easy to show that k is a constant; q_1 is rational, and α_1 does not involve θ_1. We repeat upon $\zeta/\theta_1^{q_1} = e^{w-q_1 x}$ the operations which we have just applied to ζ and thus obtain the result stated in (b) of our lemma.

It remains to prove that for $n = 1$, the function $\xi(u, v)$ which appears in (a) and (b) is a constant. We shall show that this is so when ζ is an exponential; if ζ is a logarithm, the treatment is very similar. By hypothesis

$$e^{B(u,v)} = e^{q_1 A_1(u,v)} \cdot e^{q_2 A_2(u,v)} \cdots e^{q_s A_s(u,v)} \cdot C(u, v),$$

where $B(u, v)$, $C(u, v)$ and the $A_i(u, v)$ are algebraic functions. From this we find directly that

$$B(u, v) = q_1 A_1(u, v) + q_2 A_2(u, v) + \cdots + q_s A_s(u, v) + \log C(u, v).$$

This shows that $C(u, v)$ must be a constant.

The lemma is proved.

III. Transformations of order one.

9. We now consider the problem mentioned in the introduction. We recall that a transformation on two variables of order one is defined by two elementary functions, each of order at most one, not both of order zero. We deal with a transformation T of order one. The inverse, T^{-1}, of T is assumed to be not only elementary, but also of order one. We propose to characterize completely all such transformations T. We write

$$(21) \qquad \begin{aligned} T: \quad & U = \varphi(u, v), \\ & V = \psi(u, v), \\ T^{-1}: \quad & U = f(u, v), \\ & V = g(u, v). \end{aligned}$$

In what follows, functions of one or more variables represented with capital letters (as $A(u, v)$, $F(\theta, u, v)$ etc.) will be understood to be algebraic functions of their variables.

Let $\varphi(u, v)$ be of order one. Suppose we are dealing with an expression of $\varphi(u, v)$ in which the number of exponentials plus the number of logarithmic sums written with independent coefficients is a minimum. Let $\xi(u, v)$ be such an exponential or sum. According to the lemma of § 5, we have

$$\xi(u, v) = K(\varphi, \varphi_1, \cdots, \varphi_k; u, v),$$

where $\varphi_1, \cdots, \varphi_k$, denote partial derivatives of $\varphi(u, v)$. Hence

$$\xi(f, g) = K(u, \varphi_1(f, g), \cdots, \varphi_k(f, g); f, g).$$

We shall prove that $\xi(f, g)$ is of order at most one in u and v. To do this, it will be sufficient to show that for any partial derivative of $\varphi(u, v)$, say $\varphi_j(u, v)$, the function $\varphi_j(f, g)$ is of order at most one. We have

$$\begin{aligned} \frac{\partial \varphi(f,g)}{\partial u} &= \frac{\partial \varphi(f,g)}{\partial f} \cdot \frac{\partial f}{\partial u} + \frac{\partial \varphi(f,g)}{\partial g} \cdot \frac{\partial g}{\partial u}, \\ \frac{\partial \varphi(f,g)}{\partial v} &= \frac{\partial \varphi(f,g)}{\partial f} \cdot \frac{\partial f}{\partial v} + \frac{\partial \varphi(f,g)}{\partial g} \cdot \frac{\partial g}{\partial v}. \end{aligned} \tag{22}$$

Since the transformation T^{-1} has a non-vanishing jacobian, we may solve for $\partial\varphi(f, g)/\partial f$ and $\partial\varphi(f, g)/\partial g$ rationally in terms of the remaining quantities in (22). As these are all of order one at most, the functions obtained from $\partial\varphi(u, v)/\partial u$ and $\partial\varphi(u, v)/\partial v$, replacing u by f, and v by g, are of order not exceeding one. An easy induction proves that $\varphi_j(f, g)$ is a function of order one at most. Hence $\xi(f, g)$ is of order not exceeding one.

If $\xi(u, v)$ is a logarithmic sum, let $\log H(u, v)$ be any logarithm of the sum. Then by the lemma of § 6, $\log H(f, g)$ is of order at most one. Let $\xi_1, \xi_2, \cdots, \xi_q$ be the exponentials and logarithmic sums appearing in the expression for $\varphi(u, v)$ which we have just been considering. If $\psi(u, v)$ is of order one, we find an expression for $\psi(u, v)$ which involves a minimum number of exponentials and logarithmic sums with independent coefficients. Let $\xi_{q+1}, \cdots, \xi_r$ be the exponentials and sums appearing in this expression of $\psi(u, v)$. It is clear that the ξ_i of the set $\xi_1, \xi_2, \cdots, \xi_r$ are not necessarily distinct (if ψ is algebraic, $r = q$). Let us consider the set of all monomials, $\theta_1, \theta_2, \cdots, \theta_t$, which appear in the set $\xi_1, \cdots, \xi_r$. Each $\theta_i(f, g)$ is of order at most one. Furthermore, $\varphi(u, v)$ and $\psi(u, v)$ may each be expressed as algebraic functions of the θ_i, u, and v. Suppose, for the sake of convenience, that $\theta_1, \cdots, \theta_s$ form an independent set while θ_j may be expressed algebraically in terms of $\theta_1, \cdots, \theta_s$ for $j = s+1, \cdots, t$. We may write

$$\begin{aligned} \varphi(u, v) &= G_1(\theta_1, \cdots, \theta_s;\ u, v), \\ \psi(u, v) &= G_2(\theta_1, \cdots, \theta_s;\ u, v). \end{aligned} \tag{23}$$

10. We shall first confine ourselves to the case in which there exists no relation

$$H[\varphi(u, v),\ \psi(u, v),\ u,\ v] \equiv 0. \tag{24}$$

If here we replace u by $f(u, v)$, and v by $g(u, v)$, we see that (24) implies

$$H[u,\ v,\ f(u, v),\ g(u, v)] \equiv 0.$$

We shall prove that in (23) $s = 2$. In the first place, suppose that only one monomial, θ_1, appears in (23). Consider the functions

$$\Phi = G_1(z, u, v),$$
$$\Psi = G_2(z, u, v),$$

obtained from (23) after replacing θ_1 by the independent variable z. If the partial derivative, $\partial\Phi/\partial z$, vanishes identically, we have a relation of the type (24); hence $\partial\Phi/\partial z \not\equiv 0$. This partial derivative is an algebraic function of z, u, and v. Suppose $\partial\Phi/\partial z$ vanishes at all points $(\theta_1(u, v), u, v)$. Then by Liouville's Principle it vanishes identically in z, u, and v. This discussion implies that $\theta_1(u, v)$ may be eliminated from the functions in (23) resulting in a relation of the type (24). Hence $s > 1$.

Suppose now that $s > 2$. Let $\theta(u, v)$ be an exponential $e^{A(u,v)}$ appearing in one of the functions (23). Since we exclude (24), $A(f, g)$ is a function of order one. If $\theta(f, g)$ is of order zero, $A(f, g)$ is of the form $\log B(u, v)$; if $\theta(f, g)$ is of order one, by the second lemma of § 7, $A(f, g)$ is of the form $C(u, v) + \log B(u, v)$. Hence we may always write

$$A(f, g) = C(u, v) + \log B(u, v)$$

with the understanding that $C(u, v)$ may be a constant. Similarly, if $\theta(u, v)$ of (23) is a logarithm, $\log A(u, v)$, we have

$$A(f, g) = C(u, v) \cdot e^{B(u,v)}.$$

By the *algebraic component of a monomial, θ, of order one,* we shall mean the algebraic function $A(u, v)$ whose exponential or logarithm is the monomial. Let $A_i(u, v)$ be the algebraic components of the θ_i of (23). We shall first show that the jacobian, J, of $A_1(u, v)$ and $A_2(u, v)$ does not vanish identically. If, on the contrary, $J \equiv 0$, we write

$$A_1(f, g) = K[A_2(f, g)]. \tag{25}$$

It is clear that θ_1 and θ_2 must both be exponentials or both be logarithms. Let us consider only the case in which θ_1 and θ_2 are exponentials; the logarithmic case is very similar. From (25) we have

$$C_1(u, v) + \log B_1(u, v) = K[C_2(u, v) + \log B_2(u, v)].$$

By the lemma of § 8 we may write

$$C_1(u, v) + \log B_1(u, v) = q[C_2(u, v) + \log B_2(u, v)] + L(u, v),$$

where $L(u, v)$ differs from $qC_2(u, v) - C_1(u, v)$ by a constant. Unless $L(u, v)$ is a constant, comparison with (25) yields a relation of type (24); if $L(u, v)$ is a constant, $\theta_1 = c\theta_2^q$ where c is a contant. This proves that $J \not\equiv 0$.

Let the monomial θ_3 in (23) be an exponential, $e^{A_3(u,v)}$. We may write

$$A_3(f, g) = M[A_1(f, g), A_2(f, g)]. \tag{26}$$

The functions $A_1(f, g)$ and $A_2(f, g)$ involve the monomials ζ_1 and ζ_2 respectively. If ζ_1 and ζ_2 do not form an independent set, we show by the method just outlined above that the jacobian of $A_1(u, v)$ and $A_2(u, v)$ vanishes identically. Applying the lemma of § 8, we may write (26) in the form

$$A_3(u, v) = q_1 A_1(u, v) + q_2 A_2(u, v) + c$$

with rational q_1 and q_2 (if θ_1 should be a logarithm, $q_1 = 0$). This shows that θ_1, θ_2, and θ_3 do not form an independent set. The case in which θ_3 is a logarithm is entirely similar. Hence $s = 2$.

Consider the functions

$$\Phi = G_1(z_1, z_2, u, v),$$
$$\Psi = G_2(z_1, z_2, u, v),$$

obtained from (23) by placing the variables z_1 and z_2 for θ_1 and θ_2 respectively. If the jacobian, K, of these functions with respect to z_1 and z_2 vanishes identically, the functions (23) yield a relation (24). Is it possible that K should vanish at all points $(\theta_1(u, v), \theta_2(u, v), u, v)$ without vanishing identically? If this is possible, we deduce that θ_1 and θ_2 do not form an independent set. This implies that we may solve (23) for θ_1 and θ_2 in terms of φ, ψ, u, and v. In the solution we replace u and v by algebraic functions of $A_1(u, v)$ and $A_2(u, v)$. We write the result in the form:

$$\theta_1 = P_1(\varphi, \psi, A_1, A_2),$$
$$\theta_2 = P_2(\varphi, \psi, A_1, A_2). \tag{27}$$

Let us assume first that θ_1 and θ_2 are both exponentials. From (27) we have

$$\theta_1(f, g) = P_1(u, v, C_1 + \log B_1, C_2 + \log B_2),$$
$$\theta_2(f, g) = P_2(u, v, C_1 + \log B_1, C_2 + \log B_2).$$

The functions $\theta_1(f, g)$ and $\theta_2(f, g)$ do not involve any logarithms. Since $\log B_1$ and $\log B_2$ form an independent set, it is clear that the functions P_1 and P_2 in (27) do not involve A_1 and A_2. In this case, φ and ψ are

algebraic functions of two exponential monomials. A transformation defined by two functions of this type clearly has an inverse of order one. This class of transformations is found under (a) in the introduction.

If we assume that θ_1 and θ_2 are logarithms, we prove with a similar argument that all admissible transformations may be reduced by algebraic transformations to the transformation (b) of the introduction.

We assume now that θ_1 is an exponential and θ_2 a logarithm. Making use of the lemma of § 8 we show quickly that

$$\begin{aligned} \theta_1 &= A_2^{q_1} \cdot R_1(\varphi, \psi), \\ \theta_2 &= q_2 A_1 + R_2(\varphi, \psi) \end{aligned} \tag{28}$$

where q_1 and q_2 are rational. From (28) we write

$$\begin{aligned} \theta_1^{(1-q_1 q_2)} &= R_1(\varphi, \psi) \cdot e^{q_1 R_2(\varphi, \psi)}, \\ (1 - q_1 q_2)\, \theta_2 &= R_2(\varphi, \psi) + \log [R_1(\varphi, \psi)]^{q_2}. \end{aligned}$$

Clearly, we may not have $q_1 q_2 = 1$. For all other permissible values of q_1 and q_2, the transformation defined by the functions φ and ψ in (28) has an inverse of order one. This class of transformation is found under class (c).

11. We now discuss transformations defined by the functions of (23) for which a relation (24) exists. The monomials which appear in (23) are of two types: Those monomials θ_i such that $A_i(f, g)$ is a function of order one (as before, $A_i(u, v)$ represents the algebraic component of θ_i); and the monomials θ_j such that $A_j(f, g)$ is a function of order zero. We shall show that both types cannot occur simultaneously in any transformation which we are studying.

Suppose, on the contrary, that θ_1 and θ_2 are two monomials in (23) for which $A_1(f, g)$ is a function of order one and $A_2(f, g)$ a function of order zero. We shall assume that θ_1 is an exponential; if θ_1 is a logarithm, the argument is similar. We write

$$\begin{aligned} A_1(f, g) &= C_1(u, v) + \log B_1(u, v), \\ A_2(f, g) &= C_2(u, v), \end{aligned}$$

where $B_1(u, v)$ is not a constant. A short argument shows that we can solve these relations for f and g; the result indicates that f and g involve one monomial only, viz., $\log B_1(u, v)$. The function $\log B_1(\varphi, \psi)$ is of order at most one. We treat two cases: $B_1(\varphi, \psi)$ is a function of order one; and $B_1(\varphi, \psi)$ is a function of order zero. In each case, we find that the functions φ and ψ involve only one monomial. This proves that for all monomials θ_i in (23) the functions $A_i(f, g)$ are of the same order.

We first treat the case in which the monomials θ_i in (23) are such that each $A_i(f, g)$ is of order one. We shall for the sake of convenience assume that the relation (24) actually involves $\psi(u, v)$. We may then obtain a relation

$$\psi(u, v) = K[u, v, \varphi(u, v)].$$

Let θ_1 in (23) be an exponential. We have

$$\begin{aligned} A_1(f, g) &= C_1(u, v) + \log B_1(u, v), \\ K(f, g, u) &= v. \end{aligned} \tag{29}$$

By a short argument we may show that the system (29) may be solved explicitly for f and g. These functions involve only one monomial, $\log B_1(u, v)$. The function $B_1(\varphi, \psi)$ is of order either zero or one. We consider it in conjunction with relation (24). We show quickly that $B_1(\varphi, \psi)$ is of order one, and that the functions φ and ψ contain only one monomial. If we assume θ_1 to be a logarithm, the discussion is similar.

We have seen that if the functions φ and ψ involve an exponential, the functions f and g involve a logarithm, and conversely. If in (23) we replace θ (remembering that $s = 1$) by z, the jacobian, J, with respect to u and v of the functions which result from the replacement vanishes identically. For if J does not vanish identically, it is easy to show that we may solve the system of equations for u and v, and after replacing z by θ, we find that

$$\begin{aligned} u &= F_1(\varphi, \psi, \theta), \\ v &= F_2(\varphi, \psi, \theta). \end{aligned} \tag{30}$$

If in (30) we replace u by f, and v by g, we obtain for f and g algebraic functions of u, v, and $\theta(f, g)$. If θ is an exponential, $\theta(f, g)$ cannot be a logarithm; if θ is a logarithm, $\theta(f, g)$ cannot be an exponential. We conclude that J is identically zero. The inverse of a transformation defined by the functions φ and ψ just characterized is clearly another transformation of order one—see case (d) of the introduction.

Finally, we consider the case in which every monomial θ_i, appearing in (23) is such that $A_i(f, g)$ is of order zero. If two monomials θ_1 and θ_2 are present, obviously the jacobian of $A_1(u, v)$ and $A_2(u, v)$ is identically zero. If we replace the θ_i by z_i the jacobian, K, with respect to u and v of the resulting functions cannot vanish identically. We therefore write

$$\begin{aligned} u &= H_1(\theta_1, \cdots, \theta_s; \varphi, \psi), \\ v &= H_2(\theta_1, \cdots, \theta_s; \varphi, \psi). \end{aligned}$$

If above we replace u by f, and v by g, remembering that $\theta_i(f, g) = \zeta_i(u, v)$ (ζ_i being a monomial), we obtain the functions inverse to the system formed by φ and ψ.

The construction of all transformations defined by functions of this type is accomplished as follows: Choose an arbitrary function $B_1(z_1, \cdots, z_s; u, v)$ where the variables are independent. Next, let $B_2(B_1, u, v)$ be so chosen that $\partial(B_1, B_2)/\partial(u, v) \not\equiv 0$. We now replace z_i by θ_i, where θ_i is either the exponential or logarithm of an algebraic function $A_i(u, v)$ so chosen that $\partial(B_2, A_i)/\partial(u, v) \equiv 0$. Then the transformation defined by

$$\varphi = B_1(\theta_1, \cdots, \theta_s; u, v), \qquad \psi = B_2(\varphi, u, v)$$

has an inverse of order one. This gives case (e).

COLUMBIA UNIVERSITY, NEW YORK, N. Y.

A PROBLEM IN THE SIMULTANEOUS REDUCTION OF TWO QUADRATIC FORMS IN INFINITELY MANY VARIABLES.[1]

By R. Lucile Anderson.

The problem of the simultaneous reduction of two quadratic forms in infinitely many variables, $Q(L, L)$ and $A(L, L)$, the latter of diagonal form and properly positive, has been considered by various writers. The Hilbert-Hellinger Theory[2] gives the reduction of the two forms in terms of an orthogonal system of characteristic vectors and characteristic differential vectors if the matrix Q is bounded and the matrix A is the unit matrix. If the matrix Q is completely continuous, the corresponding system of linear equations has no characteristic differential vectors not identically zero and the reduction theorems are obtained in terms of characteristic vectors alone. These characteristic vectors L_α are of Hilbert space and are solutions corresponding to the characteristic numbers λ_α of the linear equation

$$AL = \lambda QL. \tag{1}$$

That the situation is changed if we do not assume the matrix A to be properly positive is seen at once from the system of equations

$$\begin{aligned}
0 &= \lambda(-x_1 + y_2),\\
0 &= \lambda\left(-\frac{1}{4}x_2 + \frac{1}{2}y_1 + \frac{1}{2}y_3\right),\\
0 &= \lambda\left(-\frac{1}{9}x_3 + \frac{1}{3}y_2 + \frac{1}{3}y_4\right),\\
&\cdots\cdots\cdots\cdots\\
0 &= \lambda\left(-\frac{1}{n^2}x_n + \frac{1}{n}y_{n-1} + \frac{1}{n}y_{n+1}\right),\\
&\cdots\cdots\cdots\cdots\\
y_1 &= \lambda\left(\frac{1}{2}x_2\right),\\
y_2 &= \lambda\left(x_1 + \frac{1}{3}x_3\right),\\
y_3 &= \lambda\left(\frac{1}{2}x_2 + \frac{1}{4}x_4\right),\\
&\cdots\cdots\cdots\\
y_n &= \lambda\left(\frac{1}{n-1}x_{n-1} + \frac{1}{n+1}x_{n+1}\right),\\
&\cdots\cdots\cdots\cdots
\end{aligned}$$

[1] Received December 11, 1930.

[2] Hellinger-Toeplitz, *Integralgleichungen und Gleichungen mit unendlich vielen Unbekannten*, Encyklopädie der Mathematischen Wissenschaften, II C 13, p. 1577.

The matrix of the right member of these equations is completely continuous. Yet the solution of these equations reduces to the solution of the system of equations

$$\begin{aligned} y_1 &= \lambda(y_1 + y_3), \\ y_2 &= \lambda(2y_2 + y_4), \\ y_3 &= \lambda(y_1 + 2y_3 + y_5), \\ &\cdots\cdots\cdots \\ y_n &= \lambda(y_{n-2} + 2y_n + y_{n+2}), \\ &\cdots\cdots\cdots\cdots \end{aligned}$$

The matrix of the right member of this system is the square of the matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \cdot & \cdot \\ 1 & 0 & 1 & 0 & \cdot & \cdot \\ 0 & 1 & 0 & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}.$$

That is, the system of equations (1) with the matrix Q completely continuous and the matrix A not properly positive, may have a continuous spectrum.

A problem in the theory of integral equations corresponding to the case in which the principal diagonal of the matrix A has only one element equal to zero has been considered by Cairns[3] and by A. Pell.[4]

The following problem in the theory of integral equations leads to the case in which the principal diagonal of the matrix A has an infinite number of elements equal to zero. It is required to find functions $v(x)$ and $g(x)$ which satisfy the integral equation

$$v(x) = \lambda\left[g(x) + \int_a^b K(x, s)\, v(s)\, ds\right]$$

and an infinite number of conditions

$$0 = \lambda \int_a^b \psi_i(x)\, v(x)\, dx \qquad (i = 1, 2, \cdots),$$

where the kernel $K(x, s)$ is symmetric and continuous in $a \leqq s \leqq b$ and the functions $\psi_i(x)$ form a normed orthogonal system which is not complete. By means of a complete normed orthogonal system of functions $\varphi_i(x)$ the solution of this problem can be reduced to a problem in the theory of infinitely many variables of the type treated in the end of § 5, the matrix Q_1 being an orthogonal matrix and not closed.

[3] Cairns, *Die Anwendung der Integralgleichungen auf die zweite Variation bei isoperimetrischen Problemen*, Diss., Göttingen, 1907.

[4] A. Pell, Bull. Amer. Math. Soc., (2), vol. 16 (1910), pp. 412–415.

In this paper we consider the problem of the simultaneous reduction of two quadratic forms assuming the matrix A to be of diagonal form and not properly positive. Any number, finite or infinite, of the elements of the principal diagonal may be zero. We assume that each element of the principal diagonal which is not zero is greater than a constant $m>0$ and consider first the special case in which each element of the principal diagonal which is different from zero is equal to unity. The results for this special case are then extended to the case in which the elements of the principal diagonal may be greater than m (§ 9).

Throughout this paper the conditions imposed upon the matrix Q are such that the system of equations (1) has only a point spectrum. We base the reduction of the two given quadratic forms to some extent upon that of two forms, $A_2(Y, Y)$ and $Q_2(Y, Y) - W(Y, Y)$, where A_2 is the unit matrix and $Q_2 - W$ is completely continuous. It follows from the Hilbert theory[5] of completely continuous quadratic forms that

$$(2) \qquad Q_2(Y, Y) - W(Y, Y) = \sum_{\alpha} (Y_\alpha, Y)^2/\lambda_\alpha,$$

where λ_α and Y_α are characteristic numbers and corresponding characteristic vectors, respectively, of the equation

$$(3) \qquad Y = \lambda(Q_2 - W)\, Y.$$

The set of vectors Y_α is orthogonal and normed.

If we denote by $Y_{\alpha'}$ those vectors which are orthogonal to all vectors Y_α, we have the expansion of the quadratic form, $A_2(Y, Y)$ in the form

$$(4) \qquad A_2(Y, Y) = \sum_{\alpha} (Y_\alpha, Y)^2 + \sum_{\alpha'} (Y_{\alpha'}, Y)^2.$$

The right members of equations (2) and (4) are absolutely convergent. The equations may also be written as

$$(2') \quad Q_2(Y, Y) - W(Y, Y) = \sum_{\alpha} (Y_\alpha, Y)\,[Q_2(Y_\alpha, Y) - W(Y_\alpha, Y)],$$

$$(4') \quad A_2(Y, Y) = \sum_{\alpha} (Y_\alpha, Y)\,(A_2 Y_\alpha, Y) + \sum_{\alpha'} (Y_{\alpha'}, Y)\,(A_2 Y_{\alpha'}, Y),$$

respectively, where the right members are absolutely convergent. We obtain for the quadratic forms $Q(L, L)$ and $A(L, L)$ expansions in forms analogous to (2′) and (4′) expressed in terms of a biorthogonal system of vectors.

Under the conditions of each section we establish the following theorem.

THEOREM A. *The two quadratic forms $A(L, L)$ and $Q(L, L)$ associated with the equation*

$$(1) \qquad AL = \lambda QL,$$

[5] Hellinger-Toeplitz, l. c., p. 1558.

admit of expansion into absolutely convergent series[6]

$$Q(L, L) = \sum_{\varrho} (\mathfrak{M}_\varrho, L)(Q\mathfrak{L}_\varrho, L), \tag{5}$$

$$A(L, L) = \sum_{\varrho} (\mathfrak{M}_\varrho, L)(A\mathfrak{L}_\varrho, L), \tag{6}$$

where the vectors $\mathfrak{L}_\varrho$ *consist of*

i) *vectors* L_α *which are solutions of* (1) *corresponding to the characteristic numbers* $\lambda_\alpha \neq 0$, *including* $\lambda = \infty$;

ii) *vectors* $L_0^{(\beta)}$ *which are solutions of* (1) *corresponding to the characteristic number* $\lambda = 0$;

iii) *vectors* $L_\mu^{(\beta)}$ *which are solutions of the non-homogeneous equation*

$$AL = \lambda QL + QL_0^{(\beta)},$$

corresponding to $\lambda = 0$;

and where the vectors $\mathfrak{M}_\varrho$ *are so determined that the two sets of vectors* $\mathfrak{M}_\varrho$ *and* $\mathfrak{L}_\varrho$ *form a biorthogonal system.*

PART 1. Notation and general discussion.

§ 1. We write the system of linear equations associated with the matrices A and Q as

$$\begin{aligned}
0 &= \lambda (q_{0\cdot 11}\, x_1 + q_{0\cdot 12}\, x_2 + \cdots + q_{1\cdot 11}\, y_1 + q_{1\cdot 12}\, y_2 + \cdots),\\
0 &= \lambda (q_{0\cdot 21}\, x_1 + q_{0\cdot 22}\, x_2 + \cdots + q_{1\cdot 21}\, y_1 + q_{1\cdot 22}\, y_2 + \cdots),\\
&\cdots\cdots\cdots\cdots\cdots\cdots,\\
y_1 &= \lambda (q_{1\cdot 11}\, x_1 + q_{1\cdot 21}\, x_2 + \cdots + q_{2\cdot 11}\, y_1 + q_{2\cdot 12}\, y_2 + \cdots),\\
y_2 &= \lambda (q_{1\cdot 12}\, x_1 + q_{1\cdot 22}\, x_2 + \cdots + q_{2\cdot 21}\, y_1 + q_{2\cdot 22}\, y_2 + \cdots),\\
&\cdots\cdots\cdots\cdots\cdots\cdots,
\end{aligned}$$

where the number of zeros on the left is infinite unless otherwise stated.[7] It is convenient to write the matrix Q in the form

$$Q = \begin{pmatrix} q_{0\cdot ik} & q_{1\cdot ik'} \\ q_{1\cdot ki'} & q_{2\cdot i'k'} \end{pmatrix} \qquad (i, k, i', k' = 1, 2, \cdots).$$

The infinite matrices $(q_{0\cdot ik})$, $(q_{1\cdot ik'})$, $(q_{1\cdot ki'})$, and $(q_{2\cdot i'k'})$ which are minors of the matrix Q we denote by Q_0, Q_1, Q_1', and Q_2, respectively. Since Q is assumed to be symmetric, it follows that

[6] Series (5) may be written as

$$\sum_{\varrho} (\mathfrak{M}_\varrho, L)(Q\mathfrak{L}_\varrho, L) = \sum_{\alpha} (M_\alpha, L)^2/\lambda_\alpha + W_0(L, L),$$

where the first term arises from characteristic values $\lambda_\alpha \neq 0$ and the second term arises from the characteristic value $\lambda = 0$.

[7] R. G. D. Richardson in considering a problem in the Calculus of Variations is lead to a system of equations of this type for which $q_{0\cdot ik} = 0$ (all i and k). Trans. Am. Math. Soc., vol. 30 (1928), p. 162.

$$Q_0 = (q_{0\cdot ik}) \quad \text{and} \quad Q_2 = (q_{2\cdot i'k'})$$

are symmetric while $Q_1' = (q_{1\cdot ki'})$ is the transpose of $Q_1 = (q_{1\cdot ik'})$.

We denote by $Q^{(n)}$ the matrix obtained from Q by striking out the first n rows and columns

$$Q^{(n)} = \begin{pmatrix} q_{0\cdot ik} & q_{1\cdot ik'} \\ q_{1\cdot ki'} & q_{2\cdot i'k'} \end{pmatrix} \qquad \begin{pmatrix} i,\ k &= n+1, n+2, \cdots \\ i',\ k' &= 1, 2, \cdots \end{pmatrix}.$$

Similarly, $Q_0^{(n)}$ is obtained from Q_0 by striking out the first n rows and columns; $Q_1^{(n)}$ from Q_1 by striking out the first n rows and $Q_1'^{(n)}$ from Q_1' by striking out the first n columns.

We denote the minors of the matrix A corresponding to Q_0, Q_1, Q_1', and Q_2 by A_0, A_1, A_1', and A_2, respectively. It follows that

$$a_{0\cdot ik} = a_{1\cdot ik'} = a_{1\cdot ki'} = 0, \qquad \text{(all } i, k, i', k')$$

and

$$a_{2\cdot i'k'} = \begin{cases} 0, & i' \neq k', \\ 1, & i' = k'. \end{cases}$$

The unit matrix is denoted by $E = (e_{ik})$,

$$e_{ik} = \begin{cases} 0, & i \neq k, \\ 1, & i = k. \end{cases}$$

We say that a matrix B is of *type* (p) if it can be written in the form

$$B = E + P,$$

where the matrix P is symmetric and completely continuous.

The vectors (x_i) and $(y_{i'})$ are denoted by X and Y, respectively.

A vector whose elements are all zero is denoted by 0.

By the *product* of two vectors X and X' we understand the sum

$$\sum x_i x_i'$$

and denote this product by (X, X').

A vector $X \neq 0$ and of Hilbert space is said to be orthogonal to a matrix B if

$$BX = \sum b_{ik} x_k = 0 \qquad (i = 1, 2, \cdots).$$

A matrix B is *closed* if there exists no vector $X \neq 0$ and of Hilbert space which is orthogonal to B.

A matrix B is *properly positive* if

$$B(X, X) \geqq 0,$$

for all X of Hilbert space, the equality sign holding only if $X = 0$.

We reserve the square bracket for a vector of the type

$$L = [(x_i); (y_{i'})] = [X; Y] \qquad (i, i' = 1, 2, \cdots).$$

In particular if $X = 0$ or $Y = 0$, we write

$$L = [0; Y] \quad \text{or} \quad L = [X; 0].$$

By the product of two vectors L and L', we understand

$$(L, L') = (X, X') + (Y, Y').$$

Operating on a vector L with the matrix Q, we obtain a vector of the same type

$$\begin{aligned} QL &= [(\sum_k q_{0 \cdot ik} x_k + \sum_{k'} q_{1 \cdot ik'} y_{k'}); (\sum_k q_{1 \cdot ki'} x_k + \sum_{k'} q_{2 \cdot i'k'} y_{k'})] \\ &= [Q_0 X + Q_1 Y; Q_1' X + Q_2 Y]. \end{aligned}$$

A set of vectors X_α $(\alpha = 1, 2, \cdots)$ is said to be *orthogonal* and *normed* if all vectors of the set satisfy the relation

$$(X_\alpha, X_\beta) = e_{\alpha\beta}.$$

The set of vectors X_α is said to be complete if there exists no vector X' which is orthogonal to all X_α.

Two sets of vectors L_α and M_α are said to form a *biorthogonal* system if a $(1:1)$ correspondence can be set up, L_α corresponding to M_α, so that the relation

$$(L_\alpha, M_\beta) = e_{\alpha\beta} \tag{7}$$

is satisfied. A biorthogonal system is said to be complete as to L, M, respectively, if the system L_α, M_α, respectively, is complete. As a special case of a biorthogonal system we have an orthogonal system in which each vector L_α is identically equal to the corresponding vector M_α.

We denote the quadratic forms corresponding to the matrices Q and A by $Q(L, L)$ and $A(L, L)$ and note that

$$Q(L, L) = Q_0(X, X) + Q_1(X, Y) + Q_1'(Y, X) + Q_2(Y, Y). \tag{8}$$

Because of the particular form of the matrix A, we have

$$A(L, L) = A_2(Y, Y) = (Y, Y). \tag{9}$$

As a consequence of (2) and (8) and of (4) and (9), we have

$$Q(L, L) = Q_0(X, X) + Q_1(X, Y) + Q_1'(Y, X) + W(Y, Y) + \sum_\alpha (Y_\alpha, Y)^2/\lambda_\alpha, \tag{10}$$

$$A(L, L) = \sum_\alpha (Y_\alpha, Y)^2 + \sum_{\alpha'} (Y_{\alpha'}, Y)^2. \tag{11}$$

It is readily seen that, written in matrix notation, the system of equations under consideration has the form

$$(12)\qquad \begin{cases} (12a) & 0 = \lambda(Q_0 X + Q_1 Y), \\ (12b) & Y = \lambda(Q_1' X + Q_2 Y), \end{cases}$$

or, more compactly

$$(1)\qquad AL = \lambda QL.$$

By a *solution* of (1) we understand a vector

$$(13)\qquad L = [X; Y],$$

which satisfies (1) and is such that the vectors X and Y are of Hilbert space.

If corresponding to $\lambda = \lambda_\alpha$ there exists a solution $L_\alpha \neq 0$ of equation (1), such that $QL_\alpha \neq 0$ if $Y_\alpha = 0$, then λ_α and L_α are said to be a *characteristic number* and corresponding *characteristic vector*, respectively, of equation (1).

This excludes a solution of the type $[X; 0]$ such that the vector X is orthogonal to both the matrices Q_0 and Q_1' which is discussed in Part 3. Such a solution which can exist only under Assumption B (§ 8) would be a solution of equation (1) for all values of λ. Furthermore, if $\lambda \neq 0$ it follows that $Y_\alpha \neq 0$.

§ 2. The system of equations (12) has always the characteristic number $\lambda = 0$ to which there corresponds an infinite number of linearly independent characteristic vectors

$$(14)\qquad J_0^{(\beta)} = [X_0^{(\beta)}; 0],$$

where

$$X_0^{(\beta)} = (e_{\beta i}).$$

We denote by $L_0^{(\beta)}$ any linear combination of the vectors $J_0^{(\beta)}$

$$(14')\qquad L_0^{(\beta)} = \sum_\gamma c_\gamma^{(\beta)} J_0^{(\gamma)} = [C^{(\beta)}; 0].$$

We prove the following theorems concerning the characteristic numbers of (1) which are different from zero, if such exist, and the corresponding characteristic vectors.

THEOREM 1. *If the matrix Q is symmetric and bounded, all the characteristic numbers of equation* (1) *are real.*

To prove this we assume that λ_α be a characteristic number of (1) which is not real and L_α a corresponding characteristic vector, not identically zero. Let

$$\lambda_\alpha = \mu_\alpha + i\nu_\alpha.$$

and

$$L_\alpha = [X_\alpha + iX'_\alpha;\ Y_\alpha + iY'_\alpha],$$

where μ_α, ν_α, X_α, X'_α, Y_α and Y'_α are real and $Y_\alpha + iY'_\alpha \neq 0$. Since L_α is a solution of (1) corresponding to the characteristic number λ_α, we have

$$AL_\alpha = \lambda_\alpha QL_\alpha.$$

Form the product $(\bar{L}_\alpha, AL_\alpha)$, where $\bar{L}_\alpha$ is the conjugate of L_α.

$$(\bar{L}_\alpha, AL_\alpha) = \lambda_\alpha(\bar{L}_\alpha, QL_\alpha).$$

This may also be written as

$$(Y_\alpha, Y_\alpha) + (Y'_\alpha, Y'_\alpha) = (\mu_\alpha + i\nu_\alpha)(\bar{L}_\alpha, QL_\alpha).$$

The left member of this equation is real and, since $Y_\alpha + iY'_\alpha \neq 0$, is not zero. The product $(\bar{L}_\alpha, QL_\alpha)$ is real. It follows that $\nu_\alpha = 0$ and in consequence λ_α is real.

THEOREM 2. *If the matrix Q is symmetric and bounded and if $L_\alpha = [X_\alpha; Y_\alpha]$ and $L_\beta = [X_\beta; Y_\beta]$ are two linearly independent characteristic vectors of equation* (12) *corresponding to the same characteristic number $\lambda_\alpha \neq 0$, and if $Y_\alpha = kY_\beta$, then the vector*

$$L_\alpha - kL_\beta = [X_\alpha - kX_\beta; 0]$$

is orthogonal to the matrix Q.

Substituting L_α and kL_β in (12a) and subtracting the second equation from the first we obtain

$$0 = \lambda_\alpha Q_0(X_\alpha - kX_\beta) + \lambda_\alpha Q_1(Y_\alpha - kY_\beta).$$

And, since $Y_\alpha = kY_\beta$ and $\lambda_\alpha \neq 0$, it follows that

$$Q_0(X_\alpha - kX_\beta) = 0.$$

Similarly, by substituting L_α and kL_β in (12b), it can be shown that

$$Q'_1(X_\alpha - kX_\beta) = 0.$$

It follows that the vector $L_\alpha - kL_\beta$ is orthogonal to the matrix Q.

In view of this theorem we make the definition: Two linearly independent solutions of equations (12), $L_\alpha = [X_\alpha; Y_\alpha]$ and $L_\beta = [X_\beta; Y_\beta]$, corresponding to the same characteristic number $\lambda_\alpha \neq 0$, are essentially distinct if and only if $Y_\alpha \neq kY_\beta$, where k is any constant.

THEOREM 3. *If the matrix Q is symmetric and bounded, the vectors Y_α of the essentially distinct solutions $L_\alpha = [X_\alpha; Y_\alpha]$ of equation* (1), *corresponding to characteristic numbers different from zero, form a normed orthogonal system.*

The proof of this theorem follows in detail the proof of the analogous theorem concerning the solutions of the equation

$$X = \lambda BX,$$

where the matrix B is symmetric and bounded.

LEMMA 1. *If the matrix B is of type* (p) (p. 233) *and closed, and if R is the resolvent*[8] *of the matrix $-P$, then*

$$B(E+R) = (E+R)B = E. \tag{15}$$

The resolvent matrix R is symmetric and completely continuous. By definition of R

$$R = -P(E+R) = (E-B)(E+R).$$

Therefore $E = B(E+R)$. Similarly, it can be shown that $E = (E+R)B$.

LEMMA 2. *If two matrices A and B which are symmetric and bounded satisfy the relation*

$$AB = BA = E,$$

and if B is properly positive, it follows that A is properly positive.

Let

$$V = AU.$$

Then

$$BV = BAU = U.$$

If we assume that $A(U, U)$ is negative, we are lead to a contradiction. For

$$A(U, U) = B(V, V),$$

and that the right member be negative is contrary to hypothesis. Further, $A(U, U) = 0$ implies $B(V, V) = 0$. Since $B(V, V) = 0$ only if $V = 0$, it follows that $A(U, U) = 0$ only if $U = 0$.

Throughout the paper, it is assumed that the *matrix Q_0 is symmetric, bounded and of type* (p), *except in those cases in which $Q_0 = 0$, and that the matrix Q_2 is symmetric and completely continuous.* In Case A we assume that

[8] Hellinger-Toeplitz, l. c., p. 1410.

A) *there exists no vector* $X \neq 0$ *and of Hilbert space which is orthogonal to both matrices* Q_0 *and* Q_1'. If one of the matrices Q_0 and Q_1' is identically zero, this implies that the other is closed.[9] In Case B we assume that

B) *there exists a finite number of linearly independent vectors* $X_\alpha \neq 0$ *and of Hilbert space which are orthogonal to both matrices* Q_0 *and* Q_1'.[10]

PART 2. Case *A*.

§ 3. Q_0 closed.[11] Q_1 completely continuous. Since the matrix Q_0 is of type (p), equation (12a) may be written as

$$X = -(PX + Q_1 Y) \tag{16}$$

for $\lambda \neq 0$, and may be considered as a non-homogeneous equation in X. The vector $Q_1 Y$ is of Hilbert space. Since Q_0 is closed, it follows that equation (16) has a unique solution[12] given by

$$X = -(E+R) Q_1 Y. \tag{17}$$

The resolvent matrix R is symmetric and completely continuous; X is of Hilbert space and relation (15) becomes

$$Q_0(E+R) = (E+R) Q_0 = E. \tag{15.3}$$

Substituting from (17) in (12b), we obtain

$$Y = \lambda [Q_2 - Q_1'(E+R) Q_1] Y. \tag{3.3}$$

The matrix $Q_2 - Q_1'(E+R) Q_1$ is symmetric and completely continuous. From the manner in which equation (3.3) was obtained, it follows that if $[X_\alpha; Y_\alpha]$ is a characteristic vector of equations (12), corresponding to the characteristic number $\lambda_\alpha \neq 0$, then Y_α satisfies equation (3.3) for $\lambda = \lambda_\alpha$. Conversely, if λ_α be a characteristic number and Y_α a corresponding characteristic vector of equation (3.3), then λ_α is a characteristic number of equations (12) to which there corresponds the characteristic vector

[9] Since the matrix Q_1' is not symmetric, this does not imply in § 5 that there exists no vector Y which is orthogonal to the matrix Q_1.

[10] That this case may arise follows from the fact that if the matrix Q_0 is not closed there exists at least one and not more than a finite number of linearly independent vectors of Hilbert space which are orthogonal to the matrix Q_0. One or more of these vectors may be orthogonal to the matrix Q_1' as well.

[11] The results of this section can also be obtained under the assumption that the matrix Q_0 has a bounded reciprocal.

[12] Hellinger-Toeplitz, l. c., p. 1410.

(13.3) $$L_\alpha = [X_\alpha;\ Y_\alpha],$$

where

$$X_\alpha = -(E+R)\,Q_1\,Y_\alpha.$$

We determine the biorthogonal system of vectors necessary for the simultaneous reduction of the two quadratic forms as follows. Let the set of vectors $\mathfrak{L}_\varrho$ consist of the vectors L_α, defined by (13.3), and $L_0^{(\beta)} = J_0^{(\beta)}$, defined by (14). Corresponding to each vector L_α, we take

$$M_\alpha = \lambda_\alpha Q L_\alpha = [0;\ Y_\alpha].$$

It follows that

$$(L_\varrho,\ M_\alpha) = (Y_\varrho,\ Y_\alpha) = e_{\varrho\alpha}$$

and

$$(L_0^{(\beta)},\ M_\alpha) = 0.$$

Corresponding to each vector $L_0^{(\beta)}$ we determine a vector $M_0^{(\beta)}$ as a linear combination of the vectors

$$Q L_0^{(\beta)} = [Q_0\,X_0^{(\beta)};\ Q_1'\,X_0^{(\beta)}].$$

The vectors $Q L_0^{(\beta)}$ are orthogonal to all L_α, since

$$\begin{aligned}(L_\alpha,\ Q L_0^{(\beta)}) &= (X_\alpha,\ Q_0\,X_0^{(\beta)}) + (Y_\alpha,\ Q_1'\,X_0^{(\beta)})\\ &= (Q_0\,X_\alpha + Q_1\,Y_\alpha,\ X_0^{(\beta)})\\ &= 0, \qquad \text{for all } \alpha \text{ and } \beta.\end{aligned}$$

It follows that any linear combination of the vectors $Q L_0^{(\beta)}$ is orthogonal to all L_α. Let

$$M_0^{(\beta)} = \sum_\gamma c_\gamma^{(\beta)}\,Q L_0^{(\gamma)}, \qquad (\beta = 1, 2, \cdots),$$

where the constants $c_\gamma^{(\beta)}$ are to be determined so that the vectors $L_0^{(\beta)}$ and $M_0^{(\beta)}$ satisfy condition (7). The vectors $M_0^{(\beta)}$ have the form

$$\begin{aligned}M_0^{(\beta)} &= \Big[\sum_{\gamma,k} c_\gamma^{(\beta)}\,q_{0\cdot ik}\,e_{\gamma k};\ \sum_{\gamma,k} c_\gamma^{(\beta)}\,q_{1\cdot ki'}\,e_{\gamma k}\Big]\\ &= \Big[\sum_k q_{0\cdot ik}\,c_k^{(\beta)};\ \sum_k q_{1\cdot ki'}\,c_k^{(\beta)}\Big]\\ &= [Q_0\,C^{(\beta)};\ Q_1'\,C^{(\beta)}].\end{aligned}$$

From the definitions of $L_0^{(\beta)}$ and $M_0^{(\beta)}$, the vector product of (7) becomes

$$\begin{aligned}(L_0^{(\varrho)},\ M_0^{(\beta)}) &= \sum_{i,k} e_{\varrho i}\,q_{0\cdot ik}\,c_k^{(\beta)}\\ &= \sum_k q_{0\cdot \varrho k}\,c_k^{(\beta)}\end{aligned}$$

and the condition for biorthogonality may be written as

$$\sum_k q_{0\cdot\varrho k}\, c_k^{(\beta)} = e_{\varrho\beta}$$

or

$$Q_0\, C^{(\beta)} = X_0^{(\beta)}.$$

As in the solution of equation (16) the vectors $C^{(\beta)}$ of Hilbert space are uniquely determined and given by

$$C^{(\beta)} = (E+R)\, X_0^{(\beta)}.$$

If there exist vectors $Y_{\alpha'}$ which are orthogonal to all the characteristic vectors Y_α of equation (3.3), we add to the biorthogonal system already determined the vectors

$$L_{\alpha'} = [X_{\alpha'};\, Y_{\alpha'}] \qquad \text{and} \qquad M_{\alpha'} = [0;\, Y_{\alpha'}],$$

where

$$X_{\alpha'} = -(E+R)\, Q_1\, Y_{\alpha'}.$$

It is easily seen that

$$Q\, L_{\alpha'} = 0. \tag{18}$$

The two systems of vectors $\mathfrak{L}_\varrho$ and $\mathfrak{M}_\varrho$ form a biorthogonal system, complete as to $\mathfrak{L}$ and $\mathfrak{M}$.

In terms of this system,

$$\mathfrak{L}_\varrho = \begin{cases} L_\alpha = [X_\alpha;\, Y_\alpha], \\ L_{\alpha'} = [X_{\alpha'};\, Y_{\alpha'}], \\ L_0^{(\beta)} = [X_0^{(\beta)};\, 0], \end{cases} \qquad \mathfrak{M}_\varrho = \begin{cases} M_\alpha = [0;\, Y_\alpha], \\ M_{\alpha'} = [0;\, Y_{\alpha'}], \\ M_0^{(\beta)} = [Q_0(E+R)X_0^{(\beta)};\, Q_1'(E+R)X_0^{(\beta)}], \end{cases}$$

the series on the left of (5) becomes

$$\sum_\varrho (\mathfrak{M}_\varrho,\, L)\,(Q\mathfrak{L}_\varrho,\, L)$$

$$\begin{aligned} = \sum_\varrho \big[& (Q_0(E+R)X_0^{(\beta)},\, X)\,(Q_0 X_0^{(\beta)},\, X) + (Q_0(E+R)X_0^{(\beta)},\, X)\,(Q_1' X_0^{(\beta)},\, Y) \\ & + (Y_\alpha,\, Y)\,(Q_0 X_\alpha + Q_1 Y_\alpha,\, X) \quad + (Y_{\alpha'},\, Y)\,(Q_0 X_{\alpha'} + Q_1 Y_{\alpha'},\, X) \\ & + (Q_1'(E+R)X_0^{(\beta)},\, Y)\,(Q_0 X_0^{(\beta)},\, X) + (Y_\alpha,\, Y)\,(Q_1' X_\alpha + Q_2 Y_\alpha,\, Y) \\ & + (Y_{\alpha'},\, Y)\,(Q_1' X_{\alpha'} + Q_2 Y_{\alpha'},\, Y) \quad + (Q_1'(E+R)X_0^{(\beta)},\, Y)\,(Q_1' X_0^{(\beta)},\, Y)\big]. \end{aligned}$$

By virtue of (15.3), (18) and the definitions of $X_0^{(\beta)}$ and L_α this reduces to

$$\sum_\varrho (\mathfrak{M}_\varrho,\, L)\,(Q\mathfrak{L}_\varrho,\, L)$$

$$= Q_0(X,\, X) + Q_1(X,\, Y) + Q_1'(Y,\, X) + W(Y,\, Y) + \sum_\alpha (Y_\alpha,\, Y)^2/\lambda_\alpha;$$

where the matrix W has the form $Q_1(E+R)Q_1'$.

Comparison with equation (10) shows that

$$Q(L, L) = \sum_{\varrho} (\mathfrak{M}_\varrho, L)(Q\mathfrak{L}_\varrho, L)$$
$$= \sum_{\alpha} (M_\alpha, L)(QL_\alpha, L) + \sum_{\beta} (M_0^{(\beta)}, L)(QL_0^{(\beta)}, L).$$

It is easily verified that we also have

$$A(L, L) = \sum_{\varrho} (\mathfrak{M}_\varrho, L)(A\mathfrak{L}_\varrho, L);$$

and hence that Theorem A has been established under the conditions of this section.

If we assume that the principal diagonal of the matrix A contains only a finite number, m, of elements equal to zero, the indices i and k run only from 1 to m. The methods and results of this section remain applicable if we leave the conditions on Q_1 and Q_2 unchanged and assume that the matrix Q_0 is closed with respect to vectors of the form

$$X = (x_1, x_2, \cdots, x_m).$$

In this case the resulting biorthogonal system contains only m vectors of the type $L_0^{(\beta)}$.

§ 4. $Q_0 = 0$. Q_1 **closed.** $Q_1 Q_1'$ **of type** (p).[13] Since $Q_0 = 0$, it follows from Assumption A that the matrix Q_1' is closed. Equation (12a) becomes

$$\lambda Q_1 Y = 0.$$

This equation is satisfied only if $\lambda = 0$ or if $Y = 0$. Corresponding to $\lambda = 0$, equation (1) has the characteristic vectors $J_0^{(\beta)}$ defined in (14). For $\lambda \neq 0$ and $Y = 0$, equation (12b) is satisfied if and only if

$$Q_1' X = 0.$$

It follows, since Q_1' is closed that $X = 0$. That is, equation (1) has only the characteristic number $\lambda = 0$ to which there correspond the characteristic vectors $J_0^{(\beta)}$.

To determine the biorthogonal system of vectors necessary for the reduction of the two quadratic forms, we first observe that the vectors $J_0^{(\beta)}$ form an orthogonal system, and let

$$L_0^{(\beta)} = M_0^{(\beta)} = J_0^{(\beta)}.$$

[13] The results of this section can also be obtained under the assumption that the matrix $Q_1 Q_1'$ has a bounded reciprocal.

The vectors

$$L_\mu^{(\beta)} = Q L_0^{(\beta)} = [0; Q_1' X_0^{(\beta)}]$$

are orthogonal to all $M_0^{(\beta)}$.

Let

$$M_\mu^{(\beta)} = \sum_\gamma c_\gamma^{(\beta)} Q L_0^{(\gamma)} \qquad (\beta = 1, 2, \cdots),$$

where the constants $c_\gamma^{(\beta)}$ are to be determined so that the vectors $L_\mu^{(\beta)}$ and $M_\mu^{(\beta)}$ satisfy the condition for biorthogonality (7). As in § 3 this condition for biorthogonality reduces to

$$Q_1 Q_1' C^{(\beta)} = X_0^{(\beta)}.$$

The matrix $Q_1 Q_1'$ is properly positive, for

$$Q_1 Q_1'(X, X) = \sum (Q_1' X)^2,$$

and hence $Q_1 Q_1'(X, X) \geqq 0$, the equality sign holding only if $Q_1' X = 0$. Since Q_1' is closed, this is possible only if $X = 0$. $Q_1 Q_1'$ is assumed to be of type (p). It is also closed. It follows that the vectors $C^{(\beta)}$ are uniquely determined and given by

$$C^{(\beta)} = (E + R) X_0^{(\beta)},$$

where R is the resolvent of the matrix[14] $-P$, and the relation (15) becomes

$$Q_1 Q_1'(E + R) = (E + R) Q_1 Q_1' = E. \tag{15.4}$$

It may be shown as a consequence of (15.4) and the fact that Q_1 is closed that

$$Q_1'(E + R) Q_1 Q_2 = Q_2.$$

The biorthogonal system of vectors thus determined has the form

$$\mathfrak{L}_\varrho = \begin{cases} L_0^{(\beta)} = [X_0^{(\beta)}; 0], \\ L_\mu^{(\beta)} = [0; Q_1' X_0^{(\beta)}], \end{cases} \qquad \mathfrak{M}_\varrho = \begin{cases} M_0^{(\beta)} = [X_0^{(\beta)}; 0], \\ M_\mu^{(\beta)} = [0; Q_1'(E + R) X_0^{(\beta)}]. \end{cases}$$

It is easily verified that the simultaneous reduction of the two quadratic forms $Q(L, L)$ and $A(L, L)$, as stated in Theorem A, is accomplished in terms of this system of vectors.

This case does not arise if the principal diagonal of the matrix A contains only a finite number of elements equal to zero. Under this condition the matrix Q_1 composed of a finite number of rows and an infinite number of columns can not be closed.

[14] The matrix P is used as a generic term to denote a symmetric and completely continuous matrix in a matrix of type (p) (p. 233).

§ 5. $Q_0 = 0$. Q_1 **not closed.** $Q_1 Q_1'$ **of type** (p). We consider the system of equations (12) to be an equation in the vector Y, involving the vector

$$Z = \lambda X, \tag{19}$$

as an arbitrary parameter, with the condition

$$Q_1 Y = 0 \tag{12a.5}$$

on the solution Y. Operating on (12b) with the matrix Q_1, we obtain

$$Q_1 Y = \lambda Q_1 Q_2 Y + Q_1 Q_1' Z. \tag{20}$$

From (12a.5) this is to be zero. Since the matrix Q_1' is closed, it can be shown as in § 4 that the matrix $Q_1 Q_1'$ is properly positive. It is assumed to be of type (p). The vector $\lambda Q_1 Q_2 Y$ is of Hilbert space. Therefore equation (20), which may be written as

$$Z = -(PZ + \lambda Q_1 Q_2 Y),$$

has a unique solution which is given by

$$Z = -\lambda(E+R) Q_1 Q_2 Y, \tag{21}$$

where R is the resolvent of the matrix $-P$ and relation (15) becomes

$$Q_1 Q_1'(E+R) = (E+R) Q_1 Q_1' = E. \tag{15.5}$$

Substituting for Z from (21) in equation (12b), we obtain an equation involving Y alone,

$$Y = \lambda[Q_2 - Q_1'(E+R) Q_1 Q_2] Y.$$

The matrix of the right member of this equation is not necessarily symmetric. If, however, the condition (12a.5) is satisfied, we may add to this member the terms,

$$-Q_2 Q_1'(E+R) Q_1 Y,$$

without destroying the equality. The equation then becomes

$$Y = \lambda[Q_2 - Q_1'(E+R) Q_1 Q_2 - Q_2 Q_1'(E+R) Q_1] Y. \tag{3.5}$$

It is easily seen that the matrix of equation (3.5) is symmetric. It is also completely continuous since the matrices Q_2 and R are completely continuous and the matrix Q_1 is bounded. From the manner in which equation (3.5) was obtained, it follows that if $[X_\alpha; Y_\alpha]$ is a characteristic vector of equations (12) corresponding to the characteristic number λ_α, then Y_α satisfies (3.5) for $\lambda = \lambda_\alpha$. Conversely, if λ_α be a characteristic number and Y_α

a corresponding characteristic vector of equation (3.5) and if Y_α satisfies condition (12a.5), then λ_α is a characteristic number of equations (12) to which there corresponds the characteristic vector

$$L_\alpha = [X_\alpha; Y_\alpha], \tag{13.5}$$

where

$$X_\alpha = -(E+R)\,Q_1\,Q_2\,Y_\alpha.$$

To determine whether or not the condition (12a.5) is satisfied by all solutions Y_α of equation (3.5), we operate on this equation with the matrix Q_1 and obtain, by virtue of (15.5),

$$Q_1\,Y = -\lambda\,Q_1\,Q_2\,Q_1'(E+R)\,Q_1\,Y, \tag{22}$$

a homogeneous equation in the vector $Q_1\,Y$. The matrix of the right member of equation (22) is completely continuous but is not symmetric. For values of λ which are not characteristic numbers of equation (22), the equation has only the trivial solution $Q_1\,Y = 0$. That is, condition (12a.5) is satisfied.

Before considering the question of the existence of characteristic numbers and corresponding characteristic vectors of equation (22), we consider whether or not equation (3.5) may have a solution of the form

$$Y_\beta = \sum_\gamma v_{\beta\gamma}\,Q\,J_0^{(\gamma)} = Q_1'\,V_\beta, \tag{23}$$

where $v_{\beta\gamma}$ are constants, the vector V_β being of Hilbert space. Substituting from (23) in equation (3.5), we obtain by virtue of (15.5)

$$Q_1'\,V_\beta = -\lambda\,Q_1'(E+R)\,Q_1\,Q_2\,Q_1'\,V_\beta. \tag{24}$$

If the vector V_β satisfies the equation

$$V_\beta = -\lambda\,(E+R)\,Q_1\,Q_2\,Q_1'\,V_\beta, \tag{22'}$$

then equation (24) is satisfied and the vector $Q_1'\,V_\beta$ is a solution of equation (3.5). Equation (22') is the adjoint of equation (22). Therefore if μ_β be a characteristic number of one of these equations, it is also a characteristic number of the other and there exists a corresponding characteristic vector V_β of (22') such that the vector $Q_1'\,V_\beta$ is a characteristic vector of the equation (3.5) corresponding to the same characteristic number μ_β. Moreover, since $Q_1\,Q_1'$ is closed,

$$Q_1\,Y_\beta = Q_1\,Q_1'\,V_\beta \neq 0.$$

That is, $Q_1'\,V_\beta$ is a characteristic vector of equation (3.5) which does not satisfy condition (12a).

It is necessary therefore to consider equation

$$U = -\mu\, Q_1\, Q_2\, Q_1'(E+R)\, U, \tag{22}$$

and its adjoint

$$V = -\mu(E+R)\, Q_1\, Q_2\, Q_1'\, V. \tag{22'}$$

Since the matrix $Q_1\, Q_1'$ is symmetric, bounded and properly positive, it follows as a consequence of (15.5) and Lemma 2 that the matrix $(E+R)$ which is symmetric and bounded is also properly positive. Moreover, the matrix

$$(E+R)\, Q_1\, Q_2\, Q_1'\,(E+R)$$

is symmetric. We may, therefore, apply to equation (22') and its adjoint equation results which have been obtained for systems of linear equations with symmetrizable matrices.[15] In particular, we know that at least one characteristic number exists; that there corresponds to each characteristic number a finite number of characteristic vectors V_β of equation (22'). The vectors V_β are of Hilbert space and, together with vectors $V_{\beta'}$ of Hilbert space such that

$$(E+R)\, Q_1\, Q_2\, Q_1'\, V_{\beta'} = 0, \tag{25}$$

form a complete system. We denote this complete system of vectors by $\overline{V}_\beta$ and the matrix $(\overline{v}_{\beta i})$ by $\overline{V}$. The matrix $\overline{V}$ is bounded. Corresponding to each vector V_β, there exists a vector U_β, a solution of equation (22) corresponding to the characteristic number μ_β and such that

$$U_\beta = (E+R)^{-1}\, V_\beta = Q_1\, Q_1'\, V_\beta. \tag{26}$$

We set

$$U_{\beta'} = (E+R)^{-1}\, V_{\beta'} = Q_1\, Q_1'\, V_{\beta'} \tag{26'}$$

and denote by $\overline{U}_\beta$ the system of vectors consisting of all vectors U_β and $U_{\beta'}$. Since the matrix $(E+R)$ is bounded, it follows from (26) and (26') that the vectors $\overline{U}_\beta$ are of Hilbert space. It is easily verified that the vectors $\overline{U}_\beta$, $\overline{V}_\beta$ form a biorthogonal system complete as to U and V. Finally, for any vector X of Hilbert space

$$Q_1\, Q_2\, Q_1'\,(E+R)\,(X, X) = \sum_\beta (U_\beta, X)\,(V_\beta, X)/\mu_\beta \tag{27}$$

and

$$E(X, X) = \sum_\beta (\overline{U}_\beta, X)\,(\overline{V}_\beta, X),$$

[15] A. Pell, Trans. Amer. Math. Soc., vol. 20, pp. 30ff.

where the series on the right are absolutely convergent. As a consequence of (26) and (26′) this last equation may be written as

$$E(X, X) = \sum_{\beta} (\bar{V}_\beta, X)\,(Q_1 Q_1' \bar{V}_\beta, X). \tag{28}$$

It follows as a consequence of (15.5) and (25) that the vectors $Q_1' V_{\beta'}$ are orthogonal to the matrix of the right member of equation (3.5) and are therefore orthogonal to all characteristic vectors of that equation.

We have seen that each vector V_β of equation (22′) corresponding to a characteristic number μ_β furnishes a characteristic vector $Q_1' V_\beta$ of equation (3.5) corresponding to the same characteristic number μ_β which does not satisfy the condition (12a.5). We assume that, corresponding to the characteristic number μ_β, there exists a solution N_β of equation (3.5) which is not of the form $Q_1' V_\beta$. Then the vector

$$N_\beta' = N_\beta - Q_1' V_\beta\,(N_\beta, Q_1' V_\beta) \tag{29}$$

is a characteristic vector of equation (3.5) corresponding to the characteristic number μ_β and the vector

$$L_\beta = [X_\beta; N_\beta'] \tag{13′.5}$$

is a characteristic vector of equation (12) for the same characteristic number. The vector N_β' is orthogonal to all characteristic vectors of equation (3.5) different from N_β and $Q_1' V_\beta$ since N_β and $Q_1' V_\beta$ are orthogonal to all other characteristic vectors of equation (3.5). N_β' is orthogonal to $Q_1' V_\beta$, since

$$(N_\beta', Q_1' V_\beta) = (N_\beta, Q_1' V_\beta) - (N_\beta, Q_1' V_\beta) = 0.$$

It follows that

$$(N_\beta', Q_1' V_\varrho) = (V_\varrho, Q_1 N_\beta') = 0 \qquad \text{(all } \varrho\text{)}. \tag{30}$$

Equation (30) states that the vector $Q_1 N_\beta'$ is orthogonal to all vectors V_ϱ. It is also orthogonal to all vectors $V_{\varrho'}$, since N_β' is a characteristic vector of equation (3.5) and therefore

$$(N_\beta', Q_1' V_{\varrho'}) = (V_{\varrho'}, Q_1 N_\beta') = 0.$$

That is, the vector $Q_1 N_\beta'$ is orthogonal to all vectors $\bar{V}_\varrho$, a complete system. This is possible if and only if

$$Q_1 N_\beta' = 0.$$

It follows that the vector L_β defined by (13′.5) is a characteristic vector of equations (12). We conclude therefore that all characteristic vectors of equation (3.5) furnish characteristic vectors of equations (12) with the

exception of those characteristic vectors of equation (3.5) which correspond to characteristic numbers μ_β and are of the form $Q_1' V_\beta$.

To determine the biorthogonal system of vectors necessary for the simultaneous reduction of the two quadratic forms, $Q(L, L)$ and $A(L, L)$, we let the set of vectors $\mathfrak{L}_\varrho$ consist of the vectors L_α defined by (13.5) and (13'.5), the vectors $L_0^{(\beta)} = [\bar{V}_\beta; 0]$ and the vectors $L_\mu^{(\beta)} = [X_\mu^{(\beta)}; Q_1' \bar{V}_\beta]$, where

$$X_\mu^{(\beta)} = -(E+R) Q_1 Q_2 Q_1' \bar{V}_\beta = \begin{cases} V_\beta/\mu_\beta & \text{if } \bar{V}_\beta = V_\beta, \\ 0 & \text{if } \bar{V}_\beta = V_{\beta'}. \end{cases}$$

The vectors $\mathfrak{M}_\varrho$ are determined so that the condition for biorthogonality (7) is satisfied and the resulting system of vectors becomes

$$\mathfrak{L}_\varrho = \begin{cases} L_\alpha = [X_\alpha; Y_\alpha], \\ L_\mu^{(\beta)} = [X_\mu^{(\beta)}; Q_1' \bar{V}_\beta], \\ L_0^{(\beta)} = [\bar{V}_\beta; 0]; \end{cases} \qquad \mathfrak{M}_\varrho = \begin{cases} M_\alpha = [0; Y_\alpha], \\ M_\mu^{(\beta)} = [0; Q_1' \bar{V}_\beta], \\ M_0^{(\beta)} = [Q_1 Q_1' \bar{V}_\beta; Q_2 Q_1' \bar{V}_\beta]. \end{cases}$$

It is readily verified that these vectors do form a biorthogonal system in terms of which we have the simultaneous reduction of the two quadratic forms $Q(L, L)$ and $A(L, L)$ as stated in Theorem A.

If, with $Q_0 = 0$, we assume that the matrix Q_1 is not closed and has a bounded reciprocal, then the system of equations (12) can be reduced by means of a non-singular linear transformation to a system of equations

$$0 = \lambda(K_1 Y),$$
$$Y = \lambda(K_1' X + K_2 Y),$$

where the matrix K_1 is an orthogonal matrix and not closed. Existence and reduction theorems can be obtained easily for this system by the methods used in this section. The results are unchanged except that the biorthogonal system of vectors $\bar{V}_\beta$, $\bar{U}_\beta$ becomes an orthogonal system of vectors under these assumptions.

The problem treated in this and in the preceding section is a problem in the Elementary Divisor Theory. The vectors $L_\mu^{(\beta)}$ which are introduced here to obtain the reduction of the quadratic form $Q(L, L)$ correspond to the so-called "principal functions"[16] which enter into the elementary divisor theory of the general unsymmetric kernel in the theory of integral equations.

If, when $Q_0 = 0$, we assume that the principal diagonal of the matrix A contains only a finite number m of elements equal to zero, we need only impose upon the matrix Q_1 the condition that it be bounded in order to obtain the simultaneous reduction of the two given quadratic forms, since

[16] Hellinger-Toeplitz, l. c., p. 1549.

Q_1' is closed. In this case the resulting biorthogonal system contains only m vectors each of the types $L_0^{(\beta)}$ and $L_\mu^{(\beta)}$.

§ 6. **Q_0 not closed. Q_1 completely continuous.** Since the matrix Q_0 is not closed, the equation

(31) $$Q_0 X = 0$$

has at least one and not more than a finite number (p. 238) of linearly independent solutions. We assume that the equation has n such solutions, $X_a = (x_{ai})$ $(a = 1, \cdots, n)$. Then in the matrix

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots \\ \cdot & \cdot & \cdot \\ x_{n1} & x_{n2} & \cdots \end{pmatrix},$$

there must be at least one square matrix of order n whose determinant Δ is different from zero. With no loss of generality we may assume this to be

$$\Delta = \begin{vmatrix} x_{11} & \cdots & x_{1n} \\ \cdot & \cdot & \cdot \\ x_{n1} & \cdots & x_{nn} \end{vmatrix}.$$

Since the vectors X_a satisfy equation (31) and since $\Delta \neq 0$, we may solve for

$$q_{0 \cdot ik} \qquad (\text{all } k;\ i \leqq n)$$

in terms of

$$q_{0 \cdot ik} \qquad (\text{all } k;\ i > n).$$

We have as a result

(32) $$q_{0 \cdot ik} = \sum_{n+1}^{\infty} d_{ij}\, q_{0 \cdot jk} \qquad (\text{all } k;\ i \leqq n),$$

where the constants d_{ij} depend upon the elements x_{ai}.

It follows that in the system of equations

(31) $$\sum_{1}^{\infty} q_{0 \cdot ik}\, x_k = 0 \qquad (i = 1, 2, \cdots),$$

the first n equations are linearly dependent upon the remaining equations and that the equation (p. 233)

(31n) $$Q_0^{(n)} X = 0,$$

has no solution other than the trivial solution. For, if (x_i) be a solution of equation (31n), then

$$X' = \begin{pmatrix} 0 \\ x_i \end{pmatrix} \qquad \begin{matrix} (i \leqq n) \\ (i > n) \end{matrix}$$

is a solution of equation (31). Further, since equation (31) has only the n linearly independent solutions X_a, X' must be linearly dependent upon the vectors X_a. This is impossible since $\Delta \neq 0$.

If $\lambda \neq 0$, equations (12) are satisfied only if

$$Q_0 X + Q_1 Y = 0, \tag{12a.6}$$

or, since Q_0 is of type (p), if

$$X = -(PX + Q_1 Y).$$

We consider these equations for $i > n$ and write them in the form

$$x_i = -\sum_{n+1}^{\infty} p_{ik} x_k - \sum_{1}^{n} p_{ik} x_k - \sum_{1}^{\infty} q_{1\cdot ik'} y_{k'} \qquad (i > n). \tag{33}$$

It has been shown that the homogeneous system of equations

$$x_i = -\sum_{n+1}^{\infty} p_{ik} x_k \qquad (i > n) \tag{31 n}$$

has only the trivial solution. It follows that equations (33) have a unique solution[17] given by

$$x_i = -\sum_{s=n+1}^{\infty} (e_{is} + r_{is}) \left(\sum_{1}^{n} q_{0\cdot sj} x_j + \sum_{1}^{\infty} q_{1\cdot sk'} y_{k'} \right) \qquad (i > n), \tag{34}$$

where the matrix $R^{(n)} = (r_{ik})$ (i and $k > n$) is the resolvent of the matrix $-P^{(n)} = (-p_{ik})$ (i and $k > n$) and relation (15) becomes

$$Q_0^{(n)}(E^{(n)} + R^{(n)}) = (E^{(n)} + R^{(n)}) Q_0^{(n)} = E^{(n)}. \tag{15.6}$$

By virtue of (32) this may be modified to

$$\begin{cases} Q_0 (E^{(n)} + R^{(n)}) = \begin{pmatrix} d_{ik} \\ e_{ik} \end{pmatrix} & \begin{matrix} (i \leqq n,\ k > n), \\ (i > n,\ k > n), \end{matrix} \\ (E^{(n)} + R^{(n)}) Q_0 = \begin{pmatrix} d_{ki} \\ e_{ik} \end{pmatrix} & \begin{matrix} (i > n,\ k \leqq n), \\ (i > n,\ k > n), \end{matrix} \end{cases} \tag{15'.6}$$

where by $Q_0(E^{(n)} + R^{(n)})$ we understand

$$\sum_{n+1}^{\infty} q_{0\cdot is}(e_{sk} + r_{sk}) \qquad (i = 1, 2, \cdots; k > n).$$

Substituting from (34) in (12a), we obtain by virtue of (15'.6)

$$0 = \lambda \left[\sum_{1}^{n} q_{0\cdot ik} x_k - \sum_{n+1}^{\infty} d_{is} \left(\sum_{1}^{n} q_{0\cdot sj} x_j + \sum_{1}^{\infty} q_{1\cdot sk'} y_{k'} \right) + \sum_{1}^{\infty} q_{1\cdot ik'} y_{k'} \right] \ (i \leqq n),$$

$$0 = \lambda \left[\sum_{1}^{n} q_{0\cdot ik} x_k - \sum_{n+1}^{\infty} e_{is} \left(\sum_{1}^{n} q_{0\cdot sj} x_j + \sum_{1}^{\infty} q_{1\cdot sk'} y_{k'} \right) + \sum_{1}^{\infty} q_{1\cdot ik'} y_{k'} \right] \ (i > n).$$

[17] $q_{0\cdot a} = p_{ik}$ for $i \neq k$.

The right members of these equations are identically zero for $i > n$. For $i \leqq n$, the remaining equations by virtue of (32) reduce to

$$0 = \lambda \sum_1^\infty \left(q_{1 \cdot ik'} - \sum_{n+1}^\infty d_{is}\, q_{1 \cdot sk'} \right) y_{k'} \qquad (i \leqq n).$$

Substituting from (34) in equation (12b) we obtain, again by virtue of (15′.6)

$$y_{i'} = \lambda \left\{ \sum_1^n \left(q_{1 \cdot ki'} - \sum_{n+1}^\infty d_{ks}\, q_{1 \cdot si'} \right) x_k + \sum_1^\infty \left[q_{2 \cdot i'k'} - \sum_{s,\, t=n+1}^\infty q_{1 \cdot ti'}\, (e_{ts} + r_{ts})\, q_{1 \cdot sk'} \right] y_{k'} \right\}.$$

Hence, setting

$$b_{0 \cdot ik} = 0 \qquad \text{(all } i \text{ and } k\text{)},$$

$$(35) \qquad b_{1 \cdot ik'} = q_{1 \cdot ik'} - \sum_{n+1}^\infty d_{is}\, q_{1 \cdot sk'} \qquad (i \leqq n;\ k' = 1, 2, \cdots),$$

$$b_{2 \cdot i'k'} = q_{2 \cdot i'k'} - \sum_{t,\, s=n+1}^\infty q_{1 \cdot ti'}\, (e_{ts} + r_{ts})\, q_{1 \cdot sk'} \qquad (i', k' = 1, 2, \cdots),$$

and denoting the matrices $(b_{0 \cdot ik})$, $(b_{1 \cdot ik'})$, $(b_{1 \cdot ki'})$, and $(b_{2 \cdot i'k'})$ by B_0, B_1, B_1', and B_2, respectively, we reduce the system of equations (12) by the substitution from (32) to

$$(36) \quad \begin{cases} (36a) & 0 = \lambda (B_0 X + B_1 Y), \\ (36b) & Y = \lambda (B_1' X + B_2 Y), \end{cases}$$

where the matrices B_0, B_1, and B_2 are defined as follows

$$(37) \qquad \begin{aligned} B_0 &= 0, \\ B_1 &= Q_1 - Q_0 (E^{(n)} + R^{(n)})\, Q_1, \\ B_2 &= Q_2 - Q_1' (E^{(n)} + R^{(n)})\, Q_1. \end{aligned}$$

It follows from the conditions imposed upon the matrix Q that the matrices B_1 and B_2 are completely continuous and that the matrix B_2 is symmetric. We denote the matrix on the right of equation (36) by B.

The elements $b_{1 \cdot ik'}$ cannot be identically zero for any $i \leqq n$ and all k' since by virtue of (32) it would follow that there exist vectors orthogonal to both matrices Q_0 and Q_1', contrary to Assumption A. Similarly, the matrix B_1' is closed, for: assume that there exists a vector (a_i) $(i = 1, 2, \cdots, n)$ which is not identically zero and which is such that

$$\sum_1^n a_i\, b_{1 \cdot ik'} = \sum_1^n a_i \left(q_{1 \cdot ik'} - \sum_{n+1}^\infty d_{is}\, q_{1 \cdot sk'} \right) = 0 \qquad (k' = 1, 2, \cdots).$$

It follows from (32) that

$$\sum_1^n a_i \left(q_{0\cdot ik} - \sum_{n+1}^{\infty} d_{is}\, q_{0\cdot sk} \right) = 0 \qquad (k = 1, 2, \cdots)$$

and hence that there exists a vector which is orthogonal to both Q_0 and Q_1'.

The matrix B_0 is identically zero. The matrix B_1 of n rows and an infinite number of columns can not be closed. It follows that the system of equations (36) is of the type treated in § 5, the matrix A containing in its principal diagonal but a finite number n of elements which are equal to zero. We may determine therefore a biorthogonal system of vectors in terms of which the two quadratic forms $B(L', L')$ and $A(L', L')$ may be reduced according to Theorem A. From this we derive the biorthogonal system

$$\mathfrak{L}_\varrho = \begin{cases} L_\alpha = [X_\alpha;\ Y_\alpha], & \\ L_\mu^{(\beta)} = [X_\mu^{(\beta)};\ B_1'\, \overline{V}_\beta], & (\beta \leqq n), \\ L_0^{(\beta)} = [\overline{V}_\beta;\ 0], & (\beta \leqq n), \\ L_0^{(\beta)} = [X_0^{(\beta)};\ 0], & (\beta > n); \end{cases}$$

$$\mathfrak{M}_\varrho = \begin{cases} M_\alpha = [0;\ Y_\alpha], & \\ M_\mu^{(\beta)} = [0;\ B_1'\, \overline{V}_\beta], & (\beta \leqq n), \\ M_0^{(\beta)} = [B_1\, B_1'\, \overline{V}_\beta;\ B_2\, B_1'\, \overline{V}_\beta], & (\beta \leqq n), \\ M_0^{(\beta)} = [Q_0 (E^{(n)} + R^{(n)})\, X_0^{(\beta)};\ Q_1' (E^{(n)} + R^{(n)})\, X_0^{(\beta)}] & \\ \qquad + \sum_1^n g_\gamma^{(\beta)}\, M_\mu^{(\gamma)} + \sum_1^n h_\gamma^{(\beta)}\, M_0^{(\gamma)}, & (\beta > n), \end{cases}$$

where:

The vectors L_α are characteristic vectors of equation (1) corresponding to the characteristic numbers $\lambda_\alpha \neq 0$. If we write

$$B_1\, B_1' = E + T$$

and denote by S the resolvent of the matrix[18] $-T$, then

$$X_\alpha = \begin{cases} -(E+S)\, B_1\, B_2\, Y_\alpha & (i \leqq n), \\ -\sum_{s=n+1}^{\infty} (e_{is} + r_{is}) \left(\sum_1^n q_{0\cdot sj}\, x_{\alpha j} + \sum_1^{\infty} q_{1\cdot sk'}\, y_{\alpha k'} \right) & (i > n). \end{cases}$$

The vectors $L_\mu^{(\beta)}$ are solutions of the non-homogeneous equations

$$\begin{aligned} AL &= \lambda QL + BL_0^{(\beta)} \\ &= \lambda QL + QL_0^{(\beta)} + \sum_{\gamma=n+1}^{\infty} k_\gamma^{(\beta)}\, QL_0^{(\gamma)}, \qquad (\beta = 1, 2, \cdots, n), \end{aligned}$$

[18] The matrices $B_1 B_1'$, T and S are finite square matrices, each of order n.

corresponding to $\lambda = 0$. The vectors $\overline{V}_\beta$ are defined as

$$\overline{V}_\beta = \begin{pmatrix} \overline{v}_{\beta i} \\ 0 \end{pmatrix} \quad \begin{matrix} (i \leqq n), \\ (i > n), \end{matrix}$$

and include vectors V_β which satisfy the equations

(22'.6) $$V = -\mu (E+S) B_1 B_2 B_1' V$$

and

$$(V_\varrho, B_1 B_1' V_\beta) = e_{\varrho\beta},$$

and vectors $V_{\beta'}$ which are orthogonal to the matrix

$$(E+S) B_1 B_2 B_1'.$$

The vectors $X_\mu^{(\beta)}$ are defined by

$$X_\mu^{(\beta)} = \begin{pmatrix} -(E+S) B_1 B_2 B_1' \overline{V}_\beta = \begin{cases} V_\beta/\mu_\beta, & \overline{V}_\beta = V_\beta \\ 0, & \overline{V}_\beta = V_{\beta'} \end{cases} \\ 0 \end{pmatrix} \quad \begin{matrix} (i \leqq n), \\ (i > n). \end{matrix}$$

The vectors $L_0^{(\beta)}$ are characteristic vectors of equation (1) corresponding to the characteristic number $\lambda = 0$. The constants $g_\gamma^{(\beta)}$ and $h_\gamma^{(\beta)}$ have the form

$$g_\gamma^{(\beta)} = \begin{cases} -\sum_{j=1}^{n} \frac{v_{\gamma j}}{\mu_\gamma} d_{j\beta} - (V_\gamma, B_1 Q_1' (E^{(n)} + R^{(n)}) X_0^{(\beta)}), & \overline{V}_\gamma = V_\gamma, \\ -(V_{\gamma'}, B_1 Q_1' (E^{(n)} + R^{(n)}) X_0^{(\beta)}), & \overline{V}_\gamma = V_{\gamma'}, \end{cases}$$

$$h_\gamma^{(\beta)} = -\sum_{j=1}^{n} \overline{v}_{\gamma j} d_{j\beta}.$$

It is easily shown that the vectors $M_0^{(\beta)}$ $(\beta > n)$ may be written in the form

$$M_0^{(\beta)} = \left[X_0^{(\beta)};\ Q_1' (E^{(n)} + R^{(n)}) X_0^{(\beta)} + \sum_1^n g_\gamma^{(\beta)} B_1' \overline{V}_\gamma + \sum_1^n h_\gamma^{(\beta)} B_2 B_1' \overline{V}_\gamma \right] \quad (\beta > n).$$

The two quadratic forms $Q(L, L)$ and $A(L, L)$ may be reduced as stated in Theorem A in terms of this biorthogonal system of vectors.

As in § 3 and § 5, these methods and results are applicable if we assume that the principal diagonal of the matrix A contains only a finite number m of elements equal to zero. We assume that the determinant of the matrix Q_0 is of rank $n < m$; Q_0 is a finite square matrix and is therefore of type (p). The resulting biorthogonal system of vectors contains only $m - n$ vectors of the type

$$L_0^{(\beta)} = [X_0^{(\beta)}; 0].$$

§ 7. $Q_1 = 0$. Since the matrix $Q_1 = 0$ it follows from Assumption A that the matrix Q_0 is closed and that the equation

$$0 = \lambda Q_0 X, \tag{12a.7}$$

has only the trivial solution, if $\lambda \neq 0$. Since the matrix Q_2 is completely continuous, the equation

$$Y = \lambda Q_2 Y \tag{12b.7}$$

has characteristic numbers λ_α and corresponding characteristic vectors Y_α. Hence, the vector

$$L_\alpha = [0; Y_\alpha] \tag{13.7}$$

is a characteristic vector of equation (1) corresponding to the characteristic number λ_α.

We determine the biorthogonal system of vectors necessary for the simultaneous reduction of the two quadratic forms as follows. Let the set of vectors $\mathfrak{L}_\varrho$ consist of the vectors L_α defined by (13.7) and the vectors $L_0^{(\beta)} = [C^{(\beta)}; 0]$ defined by (14′). Corresponding to each vector L_α, let

$$M_\alpha = \lambda_\alpha Q L_\alpha = [0; Y_\alpha].$$

Corresponding to each vector $L_0^{(\beta)}$, let

$$M_0^{(\beta)} = Q J_0^{(\beta)} = [Q_0 X_0^{(\beta)}; 0].$$

We determine vectors $C^{(\beta)}$ so that the vectors $L_0^{(\beta)}$ and $M_0^{(\beta)}$ satisfy the condition for biorthogonality (7). As in § 3 these vectors are uniquely determined and given by

$$C^{(\beta)} = (E+R) X_0^{(\beta)},$$

where the matrix R is the resolvent of the matrix $-P$ and relation (15) becomes

$$Q_0 (E+R) = (E+R) Q_0 = E. \tag{15.7}$$

In terms of this biorthogonal system,

$$\mathfrak{L}_\varrho = \begin{cases} L_\alpha = [0; Y_\alpha], \\ L_0^{(\beta)} = [(E+R) X_0^{(\beta)}; 0], \end{cases} \qquad \mathfrak{M}_\varrho = \begin{cases} M_\alpha = [0; Y_\alpha], \\ M_0^{(\beta)} = [Q_0 X_0^{(\beta)}; 0], \end{cases}$$

we obtain the simultaneous reduction of the two quadratic forms as stated in Theorem A.

The methods and results of this section remain applicable if we assume that the principal diagonal of the matrix A contains only a finite number

m of elements which are equal to zero and that the matrix Q_0 be closed with respect to vectors of the type $X = (x_1, x_2, \cdots, x_n)$. The matrix Q_0 is a finite square matrix and is therefore of type (p). The resulting biorthogonal system contains only m vectors of the type $L_0^{(\beta)}$.

PART 3. **Case B.**

§ 8. We assume in this section that there exist r linearly independent vectors (p. 238)

$$X_a \qquad (a = 1, 2, \cdots, r),$$

not identically zero and of Hilbert space which are orthogonal to both matrices Q_0 and Q_1'. It follows at once that the vectors

$$[X_a;\ 0] \qquad (a = 1, 2, \ldots, r) \tag{38}$$

are orthogonal to the matrix Q and are solutions of equation (1) for all values of λ.

As in § 6 we may solve for the elements

$$q_{0\cdot ik} \text{ and } q_{1\cdot ik'} \qquad (i \leqq r;\ k, k' = 1, 2, \cdots),$$

obtaining

$$\begin{cases} q_{0\cdot ik} = \sum_{r+1}^{\infty} d_{ij}\, q_{0\cdot jk} & (\text{all } k;\ i \leqq r), \\ q_{1\cdot ik'} = \sum_{r+1}^{\infty} d_{ij}\, q_{1\cdot jk'} & (\text{all } k';\ i \leqq r), \end{cases} \tag{39}$$

where the constants d_{ij} depend upon the elements x_{ai} and satisfy the relation

$$\sum_{1}^{r} d_{ij}\, x_{ai} = -x_{aj} \qquad (a = 1, 2, \cdots, r;\ j = r+1, \cdots). \tag{40}$$

Substituting from (39) in equations (12), we obtain

$$\begin{cases} 0 = \lambda\left[\sum_{r+1}^{\infty} q_{0\cdot ik}\left(x_k + \sum_{1}^{r} d_{jk}\, x_j\right) + \sum_{1}^{\infty} q_{1\cdot ik'}\, y_{k'}\right] & (i = 1, 2, \cdots), \\ y_{i'} = \lambda\left[\sum_{r+1}^{\infty} q_{1\cdot ki'}\left(x_k + \sum_{1}^{r} d_{jk}\, x_j\right) + \sum_{1}^{\infty} q_{2\cdot i'k'}\, y_{k'}\right] & (i' = 1, 2, \cdots), \end{cases}$$

or setting

$$x_i' = x_i + \sum_{1}^{r} d_{ji}\, x_j \qquad (i = r+1, \cdots), \tag{41}$$

$$0 = \lambda\left[\sum_{r+1}^{\infty} q_{0\cdot ik}\, x_k' + \sum_{1}^{\infty} q_{1\cdot ik'}\, y_{k'}\right] \qquad (i = 1, 2, \cdots),$$

$$y_{i'} = \lambda\left[\sum_{r+1}^{\infty} q_{1\cdot ki'}\, x_k' + \sum_{1}^{\infty} q_{2\cdot i'k'}\, y_{k'}\right] \qquad (i' = 1, 2, \cdots).$$

Since by virtue of (39), the first r equations of this system are linearly dependent upon the remaining equations, we consider the system

$$(42)\qquad \begin{cases} (42\,a) & 0 = \lambda(Q_0^{(r)}\,X' + Q_1^{(r)}\,Y), \\ (42\,b) & Y = \lambda(Q_1'^{(r)}\,X' + Q_2\,Y). \end{cases}$$

The method which was used in § 6 to show that the matrix $Q_0^{(n)}$ is closed can be used here to show that there exists no vector which is orthogonal to both matrices $Q_0^{(r)}$ and $Q_1^{(r)}$. The system of equations (42) is, therefore, of the type treated in Case *A*.

Corresponding to each characteristic vector

$$(13'.8)\qquad L_\alpha' = [X_\alpha'; Y_\alpha]$$

of equations (42) there exist $r+1$ linearly independent characteristic vectors of equations (12),

$$L_\alpha = [X_\alpha^{(a)}; Y_\alpha], \qquad (a = 1, 2, \cdots, r),$$

where

$$X_\alpha^{(a)} = \begin{pmatrix} x_{\alpha i} & (i \leqq r), \\ x_{\alpha i} + x'_{\alpha i} & (i > r), \end{pmatrix}$$

and

$$(13.8)\qquad L_\alpha = [X_\alpha; Y_\alpha],$$

where

$$X_\alpha = \begin{pmatrix} 0 & (i \leqq r), \\ x'_{\alpha i} & (i > r). \end{pmatrix}$$

These characteristic vectors are not essentially distinct (p. 236). We take as the one characteristic vector of equations (12) corresponding to a characteristic vector of equations (42), the vector L_α defined in (13.8). Equations (42) and (12) both have as characteristic vectors corresponding to the characteristic number $\lambda = 0$, the vectors

$$(14.8)\qquad J_0^{(\beta)} = [X_0^{(\beta)}; 0] \qquad (\beta = r+1, \cdots).$$

The vectors $J_0^{(\beta)}$ $(\beta = 1, 2, \cdots, r)$ which are also solutions of equations (12) corresponding to the characteristic number $\lambda = 0$ are linearly dependent upon the vectors $J_0^{(\beta)}$ $(\beta = r+1, \cdots)$ and the vectors $[X_a; 0]$, defined in (38), and hence are not considered to be characteristic vectors of equations (12) distinct from the vectors $J_0^{(\beta)}$ $(\beta = r+1, \cdots)$.

The results obtained in Part 2 may be applied to the system of equations (42). We have therefore the simultaneous reduction of the two quadratic forms $Q(L', L')$ and $A(L', L')$ in terms of a biorthogonal system of vectors, $\mathfrak{L}_\varrho'$, $\mathfrak{M}_\varrho'$, the particular form of these vectors depending upon

the nature of the matrices $Q_0^{(r)}$ and $Q_1^{(r)}$. To obtain the simultaneous reduction of the two quadratic forms $Q(L, L)$ and $A(L, L)$, we modify this system in the following manner. From each vector $\mathfrak{L}'_\varrho$ we obtain a vector $\mathfrak{L}_\varrho$, by setting

$$x_i = 0, \qquad (i \leqq r).$$

From each vector $\mathfrak{M}'_\varrho$ we obtain a vector $\mathfrak{M}_\varrho$ by extending the definition of x_i $(i = r+1, \cdots)$ to x_i $(i = 1, 2, \cdots, r)$.

For example, if the matrix $Q_0^{(r)}$ is not closed so that the equation

$$Q_0^{(r)} X = 0$$

has $n - r$ linearly independent solutions, X_a $(a = r+1, \cdots, n)$ and if the matrix $Q_1^{(r)}$ is completely continuous, we have the simultaneous reduction of the two quadratic forms $Q^{(r)}(L', L')$ and $A^{(r)}(L', L')$ in terms of the biorthogonal system

$$\mathfrak{L}'_\varrho = \begin{cases} L'_\alpha = [X'_\alpha;\ Y_\alpha], \\ L'^{(\beta)}_\mu = [X'^{(\beta)}_\mu;\ B'^{(r)}_1 \overline{V}'_\beta], & (\beta = r+1, \cdots, n), \\ L'^{(\beta)}_0 = [\overline{V}'_\beta;\ 0], & (\beta = r+1, \cdots, n), \\ L'^{(\beta)}_0 = [X^{(\beta)}_0;\ 0], & (\beta = n+1, \cdots); \end{cases}$$

$$\mathfrak{M}'_\varrho = \begin{cases} M'_\alpha = [0;\ Y_\alpha], \\ M'^{(\beta)}_\mu = [0;\ B'^{(r)}_1 \overline{V}'_\beta], & (\beta = r+1, \cdots, n), \\ M'^{(\beta)}_0 = [B^{(r)}_1 B'^{(r)}_1 \overline{V}'_\beta;\ B_2 B'^{(r)}_1 \overline{V}'_\beta], & (\beta = r+1, \cdots, n), \\ M'^{(\beta)}_0 = [Q^{(r)}_0 (E^{(n)} + R^{(n)}) X^{(\beta)}_0;\ Q'^{(r)}_1 (E^{(n)} + R^{(n)}) X^{(\beta)}_0] \\ \qquad + \sum_{\gamma=r+1}^{n} g^{(\beta)}_\gamma M'^{(\gamma)}_\mu + \sum_{\gamma=r+1}^{n} h^{(\beta)}_\gamma M'^{(\gamma)}_0 & (\beta = n+1, \cdots), \end{cases}$$

where the vectors $\mathfrak{L}'_\varrho$ and $\mathfrak{M}'_\varrho$ are defined with reference to the system of equations (42), as in § 6. From this system we obtain the biorthogonal system

$$\mathfrak{L}_\varrho = \begin{cases} L_\alpha = [X_\alpha;\ Y_\alpha], \\ L^{(\beta)}_\mu = [X^{(\beta)}_\mu;\ B'_1 \overline{V}_\beta], & (\beta = r+1, \cdots, n), \\ L^{(\beta)}_0 = [\overline{V}_\beta;\ 0], & (\beta = r+1, \cdots, n), \\ L^{(\beta)}_0 = [X^{(\beta)}_0;\ 0], & (\beta = n+1, \cdots); \end{cases}$$

$$\mathfrak{M}_\varrho = \begin{cases} M_\alpha = [0;\ Y_\alpha], \\ M^{(\beta)}_\mu = [0;\ B'_1 \overline{V}_\beta], & (\beta = r+1, \cdots, n), \\ M^{(\beta)}_0 = [B_1 B'_1 \overline{V}_\beta;\ B_2 B'_1 \overline{V}_\beta], & (\beta = r+1, \cdots, n), \\ M^{(\beta)}_0 = [Q_0 (E^{(n)} + R^{(n)}) X^{(\beta)}_0;\ Q'_1 (E^{(n)} + R^{(n)}) X^{(\beta)}_0] \\ \qquad + \sum_{\gamma=r+1}^{n} g^{(\beta)}_\gamma M^{(\gamma)}_\mu + \sum_{\gamma=r+1}^{n} h^{(\beta)}_\gamma M^{(\gamma)}_0 & (\beta = n+1, \cdots), \end{cases}$$

where in each vector $\mathfrak{L}_\varrho$ the elements x_i $(i \leqq r)$ are zero. It is in terms of this biorthogonal system that we obtain the simultaneous reduction of the two quadratic forms $Q(L, L)$ and $A(L, L)$.

In solving the system of equations (42) we obtain

$$q_{0\cdot ik} = \sum_{n+1}^{\infty} d'_{ij}\, q_{0\cdot jk} \qquad (i = r+1, \cdots, n;\ k = 1, 2, \cdots). \tag{32.8}$$

The modified matrix product of (15) becomes

$$Q_0\,(E^{(n)} + R^{(n)}) = \begin{cases} \sum_{r+1}^{n} d_{ij}\, d'_{jk} + d_{ik} & (i \leqq r;\ k > n), \\ d'_{ik} & (r < i \leqq n;\ k > n), \\ e_{ik} & (i > n;\ k > n). \end{cases}$$

The vectors $B_1\, B'_1\, \overline{V}_\beta$ have the elements,

$$\begin{cases} \sum_{s'=1}^{\infty} \sum_{j,t=r+1}^{n} d_{ij}\, b_{1\cdot js'}\, b_{1\cdot ts'}\, \overline{v}_{\beta t} & (i \leqq r), \\ \sum_{s'=1}^{\infty} \sum_{t=r+1}^{n} b_{1\cdot is'}\, b_{1\cdot ts'}\, \overline{v}_{\beta t} & (r < i \leqq n), \\ 0 & (i > n), \end{cases}$$

and it follows that the vectors $M_0^{(\beta)}$ $(\beta = n+1, \cdots)$ may be written in the form

$$M_0^{(\beta)} = \left[X^{(\beta)};\ Q'_1(E^{(n)} + R^{(n)})\, X_0^{(\beta)} + \sum_{\gamma=r+1}^{n} g_\gamma^{(\beta)}\, B'_1\, \overline{V}_\gamma + \sum_{\gamma=r+1}^{n} h_\gamma^{(\beta)}\, B_2\, B'_1\, \overline{V}_\gamma\right],$$

where

$$X^{(\beta)} = \begin{cases} d_{i\beta} & (i \leqq r), \\ 0 & (r < i \leqq n), \\ e_{i\beta} & (i > n). \end{cases}$$

With the usual modifications these results can be applied to the case in which the principal diagonal of the matrix A contains only a finite number m of elements which are equal to zero.

PART 4.

§ 9. We assume in this section that the elements of the matrix A_2 are such that

$$a_{2\cdot i'k'} = \begin{cases} 0 & (i' \neq k') \\ a_{i'} & (i' = k'), \end{cases}$$

where $a_{i'} > m > 0$. Again we base the simultaneous reduction of the two quadratic forms $A(L, L)$ and $Q(L, L)$ upon that of two forms $A_2(Y, Y)$

and $Q_2(Y, Y) - W(Y, Y)$, where the matrix $Q_2 - W$ is symmetric and completely continuous.

There is associated with these two quadratic forms the equation

$$A_2 Y = \lambda (Q_2 - W) Y. \tag{43}$$

If we let

$$z_{i'} = \sqrt{a_{i'}}\, y_{i'}, \tag{44}$$

this equation may be written as

$$Z = \lambda S Z, \tag{45}$$

where

$$s_{i'k'} = \frac{q_{2.i'k'} - w_{i'k'}}{\sqrt{a_{i'}\, a_{k'}}}.$$

The matrix S is symmetric and completely continuous. From each characteristic vector Z_α of equation (45) corresponding to a characteristic number λ_α, we obtain a characteristic vector Y_α of equation (43) corresponding to the same characteristic number. The vectors Z_α form an orthogonal system, that is

$$(Z_\alpha, Z_\beta) = e_{\alpha\beta}, \tag{46}$$

and the quadratic form $S(Z, Z)$ admits of the expansion

$$S(Z, Z) = \sum_\alpha (Z_\alpha, Z)^2 / \lambda_\alpha. \tag{47}$$

By virtue of (44) equations (46) and (47) may be written as

$$(A_2 Y_\alpha, Y_\beta) = e_{\alpha\beta}, \tag{48}$$

$$Q_2(Y, Y) - W(Y, Y) = \sum_\alpha (A_2 Y_\alpha, Y)^2 / \lambda_\alpha, \tag{49}$$

respectively.

Similarly it can be shown that, if we denote by $Y_{\alpha'}$ those vectors which are orthogonal to all vectors $A_2 Y_\alpha$, the quadratic form $A_2(Y, Y)$ may be expanded in the form

$$A_2(Y, Y) = \sum_\alpha (A_2 Y_\alpha, Y)^2 + \sum_{\alpha'} (A_2 Y_{\alpha'}, Y)^2. \tag{50}$$

The right members of equations (49) and (50) are absolutely convergent.

Under assumptions similar to those of § 3–§ 8, we are able to determine biorthogonal systems of vectors in terms of which we obtain the simultaneous reduction of the two quadratic forms $A(L, L)$ and $Q(L, L)$.

For example, we assume that $Q_0 = 0$, that the matrix Q_1 is not closed and that the matrix $Q_1 A_2^{-1} Q_1'$ is of the form $\sqrt{A_2}\, E + P^*$, where P^* is symmetric and completely continuous. By $\sqrt{A_2}$ we understand the matrix

$$\sqrt{A_2} = (\sqrt{a_{2\cdot ik}}).$$

We consider the associated system of equations

$$(12.9)\quad \begin{cases} (12.9\,\text{a}) & 0 = \lambda(Q_0 X + Q_1 Y), \\ (12.9\,\text{b}) & A_2 Y = \lambda(Q_1' X + Q_2 Y). \end{cases}$$

Operating on (12.9b) with the matrix $Q_1 A_2^{-1}$, we obtain

$$Q_1 Y = \lambda(Q_1 A_2^{-1} Q_1' X + Q_1 A_2^{-1} Q_2 Y).$$

From (12.9a) this is to be zero. It can be shown as in § 4 that the matrix $Q_1 A_2^{-1} Q_1'$ is properly positive. It is assumed to be of type (p). The vector $\lambda Q_1 A_2^{-1} Q_2 Y$ is of Hilbert space. We can therefore solve for the vector $Z = \lambda X$, and obtain

$$(21.9)\qquad Z = -\lambda(E+R) Q_1 A_2^{-1} Q_2 Y,$$

where R is the resolvent of the matrix $-P$ and the relation (15) becomes

$$(15.9)\qquad Q_1 A_2^{-1} Q_1'(E+R) = (E+R) Q_1 A_2^{-1} Q_1' = E.$$

Substituting for Z from (21.9) in equation (12.9b), there results an equation involving Y alone,

$$A_2 Y = \lambda(Q_2 - Q_1'(E+R) Q_1 A_2^{-1} Q_2) Y,$$

which, if $Q_1 Y = 0$, may be written as

$$(43')\quad A_2 Y = \lambda(Q_2 - Q_1'(E+R) Q_1 A_2^{-1} Q_2 - Q_2 A_2^{-1} Q_1'(E+R) Q_1) Y.$$

The matrix of the right member of equation (43′) is symmetric and completely continuous.

It can be shown as in § 5 that there exist values of λ to which there correspond solutions of equation (43′) of the form $A_2^{-1} Q_1' V_\beta$. These characteristic vectors do not satisfy the condition $Q_1 Y = 0$ and hence do not furnish solutions of equations (12.9). Corresponding to every characteristic vector of equation (43′) which is not of the form $A_2^{-1} Q_1' V_\beta$ there exists a characteristic vector of equation (12.9),

$$L_\alpha = [X_\alpha;\ Y_\alpha],$$

where

$$X_\alpha = -(E+R)\,Q_1\,A_2^{-1}\,Q_1'\,Y_\alpha.$$

Finally, the two given quadratic forms may by expressed as

$$Q(L, L) = \sum_\varrho (\mathfrak{M}_\varrho, L)\,(Q\mathfrak{L}_\varrho, L)$$

and

$$A(L, L) = \sum_\varrho (\mathfrak{M}_\varrho, L)\,(A\mathfrak{L}_\varrho, L).$$

The vectors $\mathfrak{L}_\varrho$ and $\mathfrak{M}_\varrho$ form a biorthogonal system which consists of the vectors

$$\mathfrak{L}_\varrho = \begin{cases} L_\alpha = [X_\alpha;\ Y_\alpha], \\ L_\mu^{(\beta)} = [X_\mu^{(\beta)};\ A_2^{-1}\,Q_1'\,\overline{V}_\beta], \\ L_0^{(\beta)} = [\overline{V}_\beta;\ 0], \end{cases}$$

$$\mathfrak{M}_\varrho = \begin{cases} M_\alpha = [0;\ A_2\,Y_\alpha], \\ M_\mu^{(\beta)} = [0;\ A_2\,A_2^{-1}\,Q_1'\,\overline{V}_\beta], \\ M_0^{(\beta)} = [Q_1\,A_2^{-1}\,Q_1'\,\overline{V}_\beta;\ Q_2\,A_2^{-1}\,Q_1'\,\overline{V}_\beta], \end{cases}$$

where the set of vectors $\overline{V}_\beta$ includes the vectors V_β mentioned above and vectors $V_{\beta'}$ orthogonal to the matrix

$$(E+R)\,Q_1\,A_2^{-1}\,Q_2\,A_2^{-1}\,Q_1';$$

the vectors $X_\mu^{(\beta)}$ are defined by

$$X_\mu^{(\beta)} = -(E+R)\,Q_1\,A_2^{-1}\,Q_2\,A_2^{-1}\,Q_1'\,\overline{V}_\beta = \begin{cases} V_\beta/\mu_\beta, & \overline{V}_\beta = V_\beta, \\ 0, & \overline{V}_\beta = V_{\beta'}. \end{cases}$$

THE PROBLEM OF BOLZA IN THE CALCULUS OF VARIATIONS.

By Gilbert Ames Bliss.[1]

1. **Introduction.** The problem to be considered in this paper is that of finding in a class of arcs,

$$y_i = y_i(x) \qquad (i = 1, \cdots, n;\ x_1 \leqq x \leqq x_2), \tag{1.1}$$

satisfying the differential equations and end-conditions

$$\varphi_\alpha(x, y, y') = 0 \qquad (\alpha = 1, \cdots, m < n), \tag{1.2}$$

$$\psi_\mu[x_1, y(x_1), x_2, y(x_2)] = 0 \qquad (\mu = 1, \cdots, p \leqq 2n+2), \tag{1.3}$$

one which minimizes an expression of the form

$$I = G[x_1, y(x_1), x_2, y(x_2)] + \int_{x_1}^{x_2} f(x, y, y')\, dx. \tag{1.4}$$

This is a very general problem of the calculus of variations first formulated by Bolza who deduced the differential equations and transversality conditions which must be satisfied by a minimizing arc.[2] His results were later proved by the author of this paper by other methods for an equivalent but different formulation of the problem and with less restrictive assumptions.[3] For the form adopted by Bliss a further necessary condition for a minimum, corresponding to that of Jacobi for simpler problems, was deduced by Cope[4] in terms of the characteristic numbers of a boundary value problem associated with the second variation and following methods suggested by Hilbert, Mason, Cairns, Richardson, Bliss, and others.[5] In a recent paper Morse has remodeled the boundary problem method of attack, with the help of his theory of broken extremals, and has shown that the necessary conditions which he lists, when suitably strengthened, are sufficient for a minimum.[6]

Morse makes an introductory statement that sufficient conditions for a minimum for the problem of Bolza are presented for the first time in

[1] Received November 16, 1931.

[2] 8, pp. 438–439. The numbers in the footnotes refer to the bibliography at the end of this paper.

[3] 12, pp. 305, 311.

[4] 14, p. 21.

[5] 9, 4, 5, 6, 13.

[6] 17, p. 528. Hahn uses a broken extremal with a single corner, 7, p. 136. His method may have been inspired by Scheeffer, 2, pp. 536, 541 and elsewhere in the paper.

his paper. This assertion is correct so far as is known to the author of the present paper. For the problem of Lagrange, however, which is the problem described above in the special case when the function G in the expression (1.4) for I is identically zero, a sufficiency proof has been suggested by A. Mayer.[7] Mayer's discussion is incomplete in some essential details. To complete it an important theorem proved by Hahn[8] is needed, besides further auxiliary theorems. It is the purpose of the present paper to apply the methods suggested by the papers of Mayer and Hahn to deduce for the problem of Bolza a necessary condition analogous to that of Jacobi for simpler problems, and to make a sufficiency proof based upon this condition and the other usual ones.[9] An interesting feature of the method is that the analogue of Jacobi's condition can be phrased in terms of the characteristic numbers of a quadratic form involving only a finite number of variables instead of the characteristic numbers of a boundary value problem.

In Section 2 below the necessary conditions deducible from the first variation are listed. They have been proved a number of times before. In Section 3 the necessary condition of Mayer is deduced. Section 4 contains a sufficiency theorem arising out of a combination of the results of Mayer and Hahn. In Section 5 proofs are given for three auxiliary theorems used in the preceding section, including an important theorem of Hahn. In Section 6 sufficient conditions for a minimum are stated and proved with the help of the theorems of Section 4. Finally in Section 7 remarks are made concerning the theory of minima of a quadratic form whose variables satisfy certain linear equations.

2. **First necessary conditions for a minimum.** In the following pages it will be understood that $\mathfrak{R}$ is a region of points (x, y, y') in which the functions f and φ_α have continuous partial derivatives up to and including those of the third order, and which is furthermore such that the functions G and ψ_μ have the same property in the domain of points $(x_1, y_{i1}, x_2, y_{i2})$ for which the points $(x_1, y_{i1}) = [x_1, y_i(x_1)]$ and $(x_2, y_{i2}) = [x_2, y_i(x_2)]$ belong to sets (x, y, y') in $\mathfrak{R}$. In order to insure the independence of the conditions (1.2) and (1.3) it is assumed that the matrices

$$\|\varphi_{\alpha y_i'}\|, \qquad \|\psi_{\mu x_1}\ \psi_{\mu y_{i_1}}\ \psi_{\mu x_2}\ \psi_{\mu y_{i_2}}\|$$

have ranks m and p, respectively, in the domains just defined.

[7] 1, p. 114; 3, p. 445.

[8] 7, p. 129.

[9] See 7, p. 133, and 10, p. 157. The methods there described would seem more precise in conjunction with the theory of Mayer.

An *admissible set* (x, y, y') is one which is interior to $\mathfrak{R}$ and satisfies the equations $\varphi_\alpha = 0$. An *admissible arc* (1.1) is one which is continuous and consists of a finite number of sub-arcs with continuously turning tangents, and whose elements (x, y, y') are all admissible.[10]

The problem of Bolza, more precisely stated, is that of finding in the class of admissible arcs satisfying the end-conditions $\psi_\mu = 0$, one which minimizes the expression I.

In the following pages it will be understood that the two end-points of the arc E_{12} whose minimizing properties are being studied satisfy the conditions $\psi_\mu = 0$, but that no other pair of distinct or coincident points on E_{12} has this property. The problem of determining the shortest arc in the plane joining two curves illustrates the reason for this assumption.

The following property of a minimizing arc is an extension, in the sense usually attributed to Du Bois Reymond, of the necessary conditions deduced by Bolza, and is a consequence of methods used by the author of this paper for the problems of Lagrange and Mayer.[11] It has been proved explicitly by Morse and Myers.[12]

I. *For every minimizing arc E_{12} for the problem of Bolza there exist constants c_i and a function*

$$F = \lambda_0 f + \lambda_\alpha(x)\varphi_\alpha$$

such that the equations

$$F_{y_i'} = \int_{x_1}^{x} F_{y_i}\, dx + c_i, \qquad \varphi_\alpha = 0$$

hold at every point of E_{12}, and furthermore such that the end-points of E_{12} satisfy, besides the equations $\psi_\mu = 0$, the condition that

$$(F - y_i' F_{y_i'})\, dx + F_{y_i'}\, dy_i \Big|_1^2 + \lambda_0\, dG = 0 \tag{2.1}$$

for every set of differentials $dx_1, dy_{i1}, dx_2, dy_{i2}$ which satisfy the equations $d\psi_\mu = 0$. The first multiplier λ_0 is a constant, and the multipliers $\lambda_\alpha(x)$ are continuous except possibly at the values of x defining the corners of E_{12}. The elements of the set $\lambda_0, \lambda_\alpha(x)$ do not vanish simultaneously at any point of E_{12}.

The dependence of the equation (2.1) upon the equations $d\psi_\mu = 0$ implies and is implied by the vanishing of all determinants of order $p+1$ of the matrix

[10] Morse defines an arc as "differentially admissible" if it satisfies the equations $\varphi_\alpha = 0$, and as "terminally admissible" if it satisfies $\psi_\mu = 0$. It is "admissible" if it satisfies both of these conditions. See 17, p. 518. The definition used in the text above was selected because of its convenience when two problems with the same equations $\varphi_\alpha = 0$ but different end-conditions $\psi_\mu = 0$ are to be considered, as in this paper and the reference 15.

[11] 15, p. 692; 12, p. 309.

[12] 16, p. 245.

$$(2.2)\qquad \left\| \begin{matrix} \lambda_0 G_{x_1} - A_1 & \lambda_0 G_{y_{i1}} - B_{i1} & \lambda_0 G_{x_2} + A_2 & \lambda_0 G_{y_{i2}} + B_{i2} \\ \psi_{\mu x_1} & \psi_{\mu y_{i1}} & \psi_{\mu x_2} & \psi_{\mu y_{i2}} \end{matrix} \right\|,$$

where A_1, B_{i1}, A_2, B_{i2} are the values of

$$(2.3)\qquad A = F - y_i' F_{y_i'}, \qquad B = F_{y_i'}$$

at the points 1 and 2.

Besides the condition just described we have also those of Weierstrass and Clebsch, described in the following theorems, which must be satisfied as in the case when the end-points are fixed:

II. The Condition of Weierstrass. *At each element* (x, y, y', λ) *of a minimizing arc which is normal on every sub-interval*[13] *the inequality*

$$E(x, y, y', \lambda, Y') \geqq 0$$

must be satisfied for every admissible set $(x, y, Y') \neq (x, y, y')$.

III. The Condition of Clebsch. *At each element* (x, y, y', λ) *of a minimizing arc which is normal on every sub-interval the inequality*

$$F_{y_i' y_k'}(x, y, y', \lambda)\, \pi_i\, \pi_k \geqq 0$$

must be satisfied by every set $(\pi_1, \cdots, \pi_n) \neq (0, \cdots, 0)$ *which is a solution of the equations* $\varphi_{\alpha y_i'}\, \pi_i = 0$.

The properties I, II, III of a minimizing arc are the ones which can be deduced without recourse to the second variation.

3. **Further necessary conditions.** The necessary condition first deduced in this section applies to a minimizing arc without corners on which the determinant

$$(3.1)\qquad \begin{vmatrix} F_{y_i' y_k'} & \varphi_{\beta y_i'} \\ \varphi_{\alpha y_k'} & 0 \end{vmatrix}$$

is different from zero, and which is normal with respect to the problem of Bolza.[14] For a normal arc the multiplier λ_0 is always different from zero and the multipliers in the form $\lambda_0 = 1$, $\lambda_\alpha(x)$ are unique. A normal arc along which the determinant (3.1) is different from zero will be called *non-singular*.

[13] For the definition of normality here used see 15, pp. 687, 688. Graves has proved the necessity of the conditions II and III for minimizing arcs which are assumed to be normal only on the whole interval $x_1\, x_2$ in a paper read before the American Mathematical Society at Minneapolis in September, 1931.

[14] 15, p. 693. The definition of normality is the same for the problem of Lagrange and the problem of Bolza. Morse and Myers have modified the definitions used in the text above, but in form rather than content. See, 16, p. 245, and the theorems in 15, pp. 688, 693.

The functions $y_i(x)$, $y_i'(x)$, $\lambda_\alpha(x)$ belonging to a normal non-singular minimizing arc must have continuous derivatives and the arc is therefore an extremal.[15] Furthermore it is a member of a $2n$-parameter family of extremals[16]

$$\begin{aligned} y_i &= y_i(x, c_1, \cdots, c_{2n}) = y_i(x, c) \qquad (x_1 \leqq x \leqq x_2), \\ \lambda_\alpha &= \lambda_\alpha(x, c_1, \cdots, c_{2n}) = \lambda_\alpha(x, c) \end{aligned} \tag{3.2}$$

for values x, c satisfying conditions of the form

$$x_{10} \leqq x \leqq x_{20}, \qquad c_s = c_{s0} \qquad (s = 1, \cdots, 2n).$$

The functions $y_i(x, c)$, $y_{ix}(x, c)$, $\lambda_\alpha(x, c)$ have continuous partial derivatives of the first three orders in a neighborhood of the values (x, c) belonging to E_{12}, and the determinant

$$\left| \begin{matrix} \partial y_i / \partial c_s \\ \partial v_i / \partial c_s \end{matrix} \right| \qquad (i = 1, \cdots, n; s = 1, \cdots, 2n)$$

in which $v_i = F_{y_i'}[x, y(x, c), y'(x, c), \lambda(x, c)]$ is different from zero along E_{12}.

If we denote by $I(x_1, x_2, c)$, $\psi_\mu(x_1, x_2, c)$ the values of I and ψ_μ on the arc (3.2) it follows, since E_{12} is a minimizing arc, that

$$I(x_1, x_2, c) \geqq I(x_{10}, x_{20}, c_0)$$

for all sets (x_1, x_2, c) which satisfy the equations $\psi_\mu(x_1, x_2, c) = 0$. Hence the first differential of I must vanish, and the second must be positive or zero, for all values dx_1, dx_2, dc_s which satisfy the equations $d\psi_\mu = 0$. We may denote by symbols δ the differentials of a function of the variables x, c with respect to the c's alone, and by symbols d the differentials with respect to all the variables. Then we have

$$\begin{aligned} dI &= I_{x_1} dx_1 + I_{x_2} dx_2 + I_{c_s} dc_s \\ &= f\, dx\Big|_1^2 + dG + \int_{x_1}^{x_2} (f_{y_i} \delta y_i + f_{y_i'} \delta y_i')\, dx, \end{aligned}$$

and with the help of the equations of variation

$$\Phi_\alpha(x, \delta y, \delta y') = \varphi_{\alpha y_i} \delta y_i + \varphi_{\alpha y_i'} \delta y_i' = 0,$$

which hold along the extremals (3.2), and the usual integration by parts, we find

$$dI = A\, dx + B_i\, dy_i\Big|_1^2 + dG$$

[15] 15, p. 684.
[16] 15, p. 687.

where A, B_i have the values (2.3) and dy_i and δy_i are related by the equations

$$dy_i = y_i' dx + \delta y_i. \tag{3.3}$$

From the condition I it follows that $dI = 0$ for all differentials dx_1, dx_2, dc_s which satisfy the equations $d\psi_\mu = 0$.

For the computation of the second variation of I we have the relations

$$\begin{aligned} d^2 I &= A\, d^2 x + B_i\, d^2 y_i + dA\, dx + dB_i\, dy_i \Big|_1^2 + d^2 G, \\ d^2 G &= G_{x_1} d^2 x_1 + G_{y_{i1}} d^2 y_{i1} + G_{x_2} d^2 x_2 + G_{y_{i2}} d^2 y_{i2} + 2Q, \\ d^2 \psi_\mu &= \psi_{\mu x_1} d^2 x_1 + \psi_{\mu y_{i1}} d^2 y_{i1} + \psi_{\mu x_2} d^2 x_2 + \psi_{\mu y_{i2}} d^2 y_{i2} + 2Q_\mu, \end{aligned} \tag{3.4}$$

where $2Q$ is the quadratic form in dx_1, dy_{i1}, dx_2, dy_{i2} whose coefficients are the second derivatives of G, and $2Q_\mu$ is the similar form for ψ_μ Since the determinants of order $p+1$ of the matrix (2.2) all vanish on E_{12} it follows that when the last equations (3.4) along E_{12} are multiplied by suitable constants e_μ and added to the first two the second differentials all disappear and

$$d^2 I = dA\, dx + dB_i\, dy_i \Big|_1^2 + 2(Q + e_\mu Q_\mu).$$

With the help of the equations $\varphi_\alpha = 0$ and Euler's equations we find

$$\begin{aligned} dA &= F_x\, dx + F_{y_i}\, dy_i - y_i'\, dB_i, \\ dB_i &= F_{y_i}\, dx + \Omega_{\eta_i'}(x, \delta y, \delta y', \delta\lambda), \end{aligned}$$

where

$$\begin{aligned} \Omega(x, \eta, \eta', \mu) &= \omega(x, \eta, \eta') + \mu_\alpha \Phi_\alpha(x, \eta, \eta'), \\ 2\omega(x, \eta, \eta') &= F_{y_i y_k} \eta_i \eta_k + 2F_{y_i y_k'} \eta_i \eta_k' + F_{y_i' y_k'} \eta_i' \eta_k'. \end{aligned}$$

After an easy calculation with the help of equation (3.3) it follows that

$$\begin{aligned} d^2 I = (F_x + y_i' F_{y_i})\, dx^2 + 2F_{y_i}\, \delta y_i\, dx + \delta y_i\, \Omega_{\eta_i'}(x, \delta y, \delta y', \delta\lambda) \Big|_1^2 \\ + 2(Q + e_\mu Q_\mu) \end{aligned} \tag{3.5}$$

where δy_i is expressed in terms of the differentials dc_s by the formula $\delta y_i = y_{ic_s}\, dc_s$.

IV. The Condition of Mayer. *For a normal non-singular minimizing arc E_{12} without corners the quadratic form* (3.5) *must satisfy the condition $d^2 I \geqq 0$ for all sets (dx_1, dx_2, dc_s) which satisfy the equations $d\psi_\mu = 0$ but whose elements are not all zero. If E_{12} is normal on every sub-interval then there can be no point conjugate to* 1 *on E_{12} between* 1 *and* 2.

The last statement of the theorem is a consequence of the theory of the Lagrange problem with fixed end-points since E_{12} must certainly minimize the integral in the expression (1.4) for I in the class of admissible arcs satisfying the equations $\varphi_\alpha = 0$ and joining the ends 1 and 2 of E_{12}.

4. **A fundamental sufficiency theorem.** For the statement of the theorems of this section the closed ε-neighborhood of the set (x_{10}, x_{20}, c_{s0}) will be denoted by $(x_{10}, x_{20}, c_{s0})_\varepsilon$. The principal theorem to be proved was suggested by the papers of Mayer and Hahn referred to above and is as follows:

THEOREM 4.1. A FUNDAMENTAL SUFFICIENCY THEOREM. *Let E_0 be an extremal arc with the following properties:*

(α) *E_0 satisfies the sufficient conditions for a proper strong relative minimum with respect to admissible arcs joining its ends;*

(β) *Its end-points satisfy the conditions $\psi_\mu = 0$ and no other pair of distinct or coincident points on E_0 has this property;*

(γ) *In the $2n$-parameter family of extremals (3.2) containing E_0 for parameter values x_{10}, x_{20}, c_{s0} the inequality $I(E) > I(E_0)$ holds for every extremal arc E satisfying the end-conditions $\psi_\mu = 0$, having its parameter values (x_1, x_2, c_s) in a neighborhood $(x_{10}, x_{20}, c_{s0})_\varepsilon$, and not identical with E_0.*

Then there exists a neighborhood $\mathfrak{F}$ of E_0 in xy-space such that $I(C) > I(E_0)$ for every admissible arc C in $\mathfrak{F}$ with ends satisfying the conditions $\psi_\mu = 0$ and not identical with E_0.

The proof of this theorem depends upon the three following auxiliary theorems which will be proved in the next section.

LEMMA 1. (Modification of Hahn's theorem). *The property (α) for E_0 implies that there exists a neighborhood $\mathfrak{F}_1$ of E_0 and a constant $\varepsilon > 0$ such that every extremal arc E of the family (3.2) with parameter values (x_1, x_2, c_s) in $(x_{10}, x_{20}, c_{s0})_\varepsilon$ minimizes the expression I in the class of admissible arcs C in $\mathfrak{F}_1$ joining the ends of E.*

LEMMA 2. *The property (α) for E_0 implies that for every $\varepsilon > 0$ there exists a $\delta > 0$ such that through each pair of end-points in the δ-neighborhoods of the ends of E_0 there passes an extremal E with parameter values (x_1, x_2, c_s) in $(x_{10}, x_{20}, c_{s0})_\varepsilon$.*

LEMMA 3. *The property (β) implies that for every $\delta > 0$ there exists a neighborhood $\mathfrak{F}$ of E_0 such that every pair of points in $\mathfrak{F}$ satisfying the conditions $\psi_\mu = 0$ necessarily has one of the points of the pair in each of the two δ-neighborhoods of the ends of E_0.*

With the help of these lemmas the proof of the theorem is as follows. Let ε be a constant effective as in the hypothesis (γ), and also as in Lemma 1 with a certain neighborhood $\mathfrak{F}_1$. Let δ be related to ε as des-

cribed in Lemma 2. Finally let $\mathfrak{F}$ be a neighborhood of E_0 interior to $\mathfrak{F}_1$ and related to δ as in Lemma 3. Then, by Lemma 3, every admissible arc C in $\mathfrak{F}$ satisfying the conditions $\psi_\mu = 0$ must have its ends in δ-neighborhoods of the ends of E_0, and according to Lemma 2 these ends will be joined by an extremal arc E with parameter values (x_1, x_2, c_s) in $(x_{10}, x_{20}, c_{s0})_\varepsilon$. We have then $I(C) \geqq I(E)$, by Lemma 1, and $I(E) \geqq I(E_0)$, by the property (γ); and hence $I(C) \geqq I(E_0)$. The equality sign can hold only if C is coincident with E_0. The theorem is therefore proved.

5. **Proofs of the three lemmas.** To prove Lemma 1 it may first be remarked that the property (α) implies the existence of a field $\mathfrak{F}_1$ having E_0 as one of its extremals.[17] The value of the invariant integral

$$I^* = \int \{(F - p_i F_{y_i'})\, dx + F_{y_i'}\, dy_i\}$$

taken from a fixed point to a variable point (x, y) of the field is a function $W(x, y)$ which has continuous partial derivatives of at least the first and second orders. At a point (x, y) of the field the canonical variables $v_i = F_{y_i'}$ for an extremal of the field have the values $v_i(x, y) = W_{y_i}$.

A further consequence of the hypothesis (α) is that the determinant (3.1) is different from zero along E_0[18] and E_0 is therefore a member, for values $(x_{10}, x_{20}, a_{i0}, b_{i0})$, of a $2n$-parameter family of extremals[19] whose equations in terms of canonical variables x, y, v have the form

$$y_i = y_i(x, a, b), \qquad v_i = v_i(x, a, b).$$

For the constants a_i, b_i we may take the initial values of the variables y_i, v_i at $x = x_{10}$. The functions y_i, v_i, y_{ix}, v_{ix} have continuous partial derivatives up to and including those of the second order at least, in a neighborhood of the values (x, a, b) belonging to E_0 when the functions f and φ_α have the continuity properties prescribed in Section 2.

The extremals of the field $\mathfrak{F}_1$ in which E_0 is imbedded have the equations

$$y_i = y_i(x, a, W_a), \qquad v_i = v_i(x, a, W_a), \tag{5.1}$$

where the symbol W_a stands for the set of derivatives $W_{a_i}(x_{10}, a)$ which are the initial values at $x = x_{10}$ of the variables v_i on the extremals of the field. For each set of sufficiently small constants β_i the equations

$$y_i = y_i(x, a, W_a + \beta), \qquad v_i = v_i(x, a, W_a + \beta), \tag{5.2}$$

[17] 15, pp. 735-736, and also 18, especially pp. 574, 576.

[18] 15, p. 735.

[19] 15, p. 687.

define a family of extremals depending upon n parameters a_i with the initial values $y_i = a_i$, $v_i = W_{a_i} + \beta_i$ at $x = x_{10}$. This family forms a field in every region of xy-space which it simply covers since the Hilbert integral belonging to the family is independent of the path in the subspace $x = x_{10}$, where it has the form

$$I^* = \int (W_{a_i} + \beta_i)\, da_i = \int d(W + \beta_i a_i),$$

and hence is independent of the path in every region simply covered by the extremals (5.2).[20] If $\mathfrak{F}_1$ is a closed field containing E_0 in its interior, simply covered by the extremals (5.1), and such that the functional determinant of the functions $y_i(x, a, W_a)$ with respect to the parameters a_k is different from zero in it, then the first n equations (5.1) have solutions $a_i = a_i(x, y)$ in $\mathfrak{F}_1$ and equations (5.2) have as initial solutions the sets (x, y, a, β) for which (x, y) is in $\mathfrak{F}_1$, $a_i = a_i(x, y)$, $\beta_i = 0$. According to well-known theorems concerning implicit functions there will therefore exist a constant $\varrho > 0$ such that for each set (x, y, β) satisfying the conditions

$$(5.3) \qquad (x, y) \text{ in } \mathfrak{F}_1, \qquad |\beta_i| \leqq \varrho$$

the equations (5.2) have solutions $a_i(x, y, \beta)$, and the functions $a_i(x, y, \beta)$ so defined are continuous and have continuous derivatives and reduce to the solutions $a_i = a_i(x, y)$ of equations (5.1) for $\beta_i = 0$. Furthermore these solutions are isolated in the sense that there exists a constant σ such that equations (5.2) have no other solutions (x, y, a, β) satisfying the conditions (5.3) and the inequalities

$$(5.4) \qquad 0 < |a_i - a_i(x, y, \beta)| \leqq \sigma.$$

It is evident from the preceding arguments that for every set β_i satisfying the conditions $|\beta_i| \leqq \varrho$ the region $\mathfrak{F}_1$ is a field with the slope-functions and multipliers

$$p_i(x, y, \beta) = y_{ix}(x, a, W_a + \beta),$$
$$l_\alpha(x, y, \beta) = \lambda_\alpha\ (x, a, W_a + \beta), \qquad [a_i = a_i(x, y, \beta)],$$

belonging to the family (5.2). Since by the hypothesis (α) the function $E(x, y, y', \lambda, Y')$ is positive for every set (x, y, y', λ) in a neighborhood $\mathfrak{N}$ of those belonging to E_0 and every Y' such that (x, y, Y') is admissible and distinct from (x, y, y'),[21] it follows that by taking $\mathfrak{F}_1$ and ϱ sufficiently small the analogous condition

[20] 15, p. 733.

[21] 15, p. 734. The condition there represented by $\Pi_{\mathfrak{N}}$ will be designated in this paper by $\Pi'_{\mathfrak{N}}$.

$$E[x, y, p(x, y, \beta), l(x, y, \beta), Y'] > 0$$
$$[(x, y, Y') \text{ admissible and } \neq (x, y, p)]$$

will hold for all sets (x, y, β) with (x, y) in $\mathfrak{F}_1$ and $|\beta_i| \leqq \varrho$. Hence for every extremal arc E of one of the fields laid on $\mathfrak{F}_1$, as described above, the inequality $I(C) > I(E)$ will hold for every admissible arc C in $\mathfrak{F}_1$ joining the ends of E but not identical with E.[22]

It remains to show only that for a sufficiently small positive constant ε every extremal arc E of the family (5.2) defined by values (x_1, x_2, a_i, β_i) in $(x_{10}, x_{20}, a_{i0}, 0)_\varepsilon$ is surely an extremal of one of the fields over $\mathfrak{F}_1$. This is evident provided that ε is taken so small that E is in $\mathfrak{F}_1$ and that along E the inequalities

$$|a_i - a_{i0}| < \sigma/2, \quad |a_{i0} - a_i(x, y, \beta)| < \sigma/2$$

hold. For then at every point of E the solution (x, y, a, β) of equations (5.2) belonging to E has

$$|a_i - a_i(x, y, \beta)| < \sigma.$$

Hence $a_i = a_i(x, y, \beta)$, since equations (5.2) have no solutions (x, y, a, β) satisfying the inequalities (5.4) except those for which $a_i = a_i(x, y, \beta)$. E is therefore surely an extremal of one of the fields. The constants a_i, β_i are the constants c_s $(s = 1, \cdots, 2n)$ of Lemma 1.

The proof of Lemma 2 is a consequence of existence theorems for implicit functions applied to the equations

$$y_{i1} = y_i(x_1, c), \quad y_{i2} = y_i(x_2, c).$$

These equations have as an initial solution the end-points and constants c_{s0} of the arc E_0, and at this initial solution the functional determinant

$$\begin{vmatrix} y_{ic_s}(x_1, c) \\ y_{ic_s}(x_2, c) \end{vmatrix}$$

is different from zero, since according to the hypothesis (α) the end-points of E_0 are not conjugate to each other. Hence the existence theorems imply the conclusion of Lemma 2.

We may prove Lemma 3 by assuming it not true and finding a contradiction. Let δ be a constant for which there is no neighborhood $\mathfrak{F}$ as in the lemma. Then in every ϱ-neighborhood of E_0 for which ϱ has

[22] 15, pp. 731–732.

a value of the form $1/2^k$, with k a positive integer, there exists a pair of points $(x, y)_k$ and $(x', y')_k$ satisfying the equations $\psi_\mu = 0$ but not both in the δ-neighborhoods of the ends of E_0. The sequence of sets $(x, y, x', y')_k$ has an accumulation set (ξ, η, ξ', η') which satisfies the equations $\psi_\mu = 0$ and has both of the points (ξ, η) and (ξ', η') on E_0. But according to the hypothesis (β) these points must then be the ends of E_0, and (ξ, η, ξ', η') can not be an accumulation point for the sequence $(x, y, x', y')_k$ since no element of this sequence has both of its points $(x, y)_k$ and $(x', y')_k$ in the δ-neighborhoods of (ξ, η) and (ξ', η').

6. **Sufficient conditions for relative minima.** In the following paragraphs the notations II′, III′, IV′ will be used to denote the conditions II, III, IV of the preceding sections strengthened to exclude the equality sign, in accordance with a precedent set by Bolza. A normal extremal arc E_{12} will be said to satisfy the condition $\text{II}'_{\mathfrak{N}}$ if the inequality

$$E(x, y, y', \lambda, Y') > 0$$

holds for every element (x, y, y', λ) in a neighborhood $\mathfrak{N}$ of those on E_{12} and for every set Y'_i such that (x, y, Y') is admissible and distinct from (x, y, y').

THEOREM 6.1. SUFFICIENT CONDITIONS FOR A STRONG RELATIVE MINIMUM. *If an admissible arc E_0 without corners is normal for the problem of Bolza, normal on every sub-interval, and satisfies the conditions* I, $\text{II}'_{\mathfrak{N}}$, III′, IV′, *then there exists a neighborhood $\mathfrak{F}$ of E_0 in xy-space such that the inequality $I(C) > I(E_0)$ holds for every admissible arc C in $\mathfrak{F}$ with ends satisfying the conditions $\psi_\mu = 0$ and not identical with E_0.*

The proof can be made by showing that the hypotheses of Theorem 6.1 imply those of Theorem 4.1. It is in the first place readily seen that the properties required for E_0 in the theorem above imply the hypothesis (α) of Theorem 4.1 provided that we can show that the condition IV′ prevents the ends of E_0 from being conjugate to each other.[23] If the ends of E_0 were conjugate the constants dc_s in the expressions $\delta y_i = y_{ic_s} dc_s$ could be selected not all zero so that the differentials δy_i would all vanish at the ends of E_0. If we should then take with these constants dc_s the values $dx_1 = dx_2 = 0$ the equations $d\psi_\mu = 0$ would be satisfied and the expression (3.5) for d^2I would vanish, which would contradict condition IV′.

The hypothesis (β) is one of the original hypotheses on E_0 in Section 2.

[23] 15, p. 735. Morse has recently proved in 18, pp. 574–576, that the assumption that E_0 is normal on every sub-interval can replace the assumption that E_0 has an extension normal on every sub-intervals in the sufficient conditions for a minimum for the Lagrange problem with fixed end-points.

Hypothesis (γ) is a consequence of the conditions I, III′ and IV′. For conditions I, III′ imply the existence of the family of extremals (3.2) in which E_0 is imbedded,[24] and from I and IV′ it follows that $dI = 0$ and $d^2I > 0$ for every set of differentials dx_1, dx_2, dc_s not all zero which satisfy the equations $d\psi_\mu = 0$. But these are conditions which from the theory of maxima and minima of functions of several variables are known to insure the inequality $I(x_1, x_2, c_s) > I(x_1, x_2, c_{s0})$ for all sets $(x_1, x_2, c_s) \neq (x_{10}, x_{20}, c_{s0})$ satisfying the equations $\psi_\mu(x_1, x_2, c_s) = 0$ and lying in a sufficiently small neighborhood $(x_{10}, x_{20}, c_{s0})_\varepsilon$.

If a normal extremal arc E for the problem of Bolza satisfies the condition III′ then there exists an ε-neighborhood $\mathfrak{N}$ of the sets (x, y, y', λ) belonging to E_0 such that the condition

$$E(x, y, y', \lambda, Y') > 0$$

holds for all sets (x, y, y', λ) in $\mathfrak{N}$ and all admissible sets (x, y, Y') which satisfy the inequalities defining $\mathfrak{N}$ and are distinct from (x, y, y').[25] Hence if we take in place of the original region $\mathfrak{N}$ a sufficiently small neighborhood $\mathfrak{N}_1$ of the sets (x, y, y') on E_0 the condition $\text{II}'_{\mathfrak{N}}$ will be satisfied for $\mathfrak{N}_1$. The preceding theorems justify then the following one:

THEOREM 6.2. SUFFICIENT CONDITIONS FOR A WEAK RELATIVE MINIMUM. *If an admissible arc E_0 without corners is normal for the problem of Bolza, normal on every sub-interval, and satisfies the conditions* I, III′, IV′ *then there exists a neighborhood $\mathfrak{N}_1$ of the sets (x, y, y') on E_0 such that the inequality $I(C) > I(E_0)$ holds for every admissible arc C in $\mathfrak{N}_1$ with ends satisfying the conditions $\psi_\mu = 0$ and not identical with E_0.*

7. Criteria for a quadratic form to be positive in a linear space. The condition of Mayer in Section 3 above required a quadratic form d^2I in the differentials dx_1, dx_2, dc_s to be positive or zero for all values of these differentials satisfying the linear equations $d\psi_\mu = 0$. The problem of determining the conditions under which this condition is satisfied is a well-known problem of algebra whose results will be briefly indicated here.

We shall consider the conditions under which a quadratic form

$$2Q(x) = a_{ik}\, x_i\, x_k \qquad (i, k = 1, \cdots, n;\ a_{ik} = a_{ki})$$

will be positive or zero in the set of points x satisfying the linear equations

$$L_\mu(x) = b_{\mu k}\, x_k = 0 \qquad (\mu = 1, \cdots, p < n) \tag{7.1}$$

[24] 15, pp. 735, 687.

[25] An argument like that of 15, pp. 736–737 justifies this statement.

in which the matrix $\|b_{\mu k}\|$ is of rank p. These conditions are the same as those which will insure the inequality $Q(x) \geqq 0$ in the set of points x satisfying

$$L_\mu(x) = 0, \qquad x_k x_k = 1. \tag{7.2}$$

The coefficients a_{ik}, $b_{\mu k}$ are all of course real.

It is well known[26] that the roots of the determinant

$$D(\sigma) = \begin{vmatrix} a_{ik} - \delta_{ik}\sigma & b_{\nu i} \\ b_{\mu k} & 0 \end{vmatrix} \qquad (\delta_{ii} = 1,\ \delta_{ik} = 0 \text{ if } i \neq k)$$

are all real, and that the minimum of $2Q$ in the set of points x satisfying the conditions (7.2) is the smallest root σ_1 of $D(\sigma)$. Hence a necessary condition for $Q(x)$ to be positive or zero in the set of points x satisfying the equations (7.1) is $\sigma_1 \geqq 0$, and furthermore if $\sigma_1 > 0$ we shall have $Q(x) > 0$ for all sets not identically zero satisfying the equations (7.1).

In many of the preceding papers on the calculus of variations the necessary condition analogous to that of Jacobi has required an inequality $\sigma_1 \geqq 0$ on the smallest characteristic number σ_1 of a boundary value problem involving differential equations and end-conditions. From the preceding paragraphs it is evident that the methods of Mayer and Hahn enable us to replace this boundary value problem by a much simpler algebraic problem of an analogous character.

BIBLIOGRAPHY.

1. A. Mayer, Zur Aufstellung der Kriterien des Maximums und Minimums der einfachen Integrale bei variabeln Grenzwerten, Leipziger Berichte, vol. 36 (1884), pp. 99–128.

2. Scheeffer, Die Maxima und Minima der einfachen Integrale zwischen festen Grenzen, Mathematische Annalen, vol. 25 (1885), pp. 522–593.

3. A. Mayer, Zur Aufstellung der Kriterien des Maximums und Minimums der einfachen Integrale bei variabeln Grenzwerten, Leipziger Berichte, vol. 48 (1896), pp. 436–465.

4. Mason, Randwertaufgaben bei gewöhnlichen Differentialgleichungen, Dissertation, Göttingen (1903). See also Mathematische Annalen, vol. 58 (1904), pp. 528–544.

5. Cairns, Die Anwendung der Integralgleichungen auf die zweite Variation bei isoperimetrischen Problemen, Dissertation, Göttingen (1907).

6. Richardson, Das Jacobische Kriterium der Variationsrechnung und die Oszillationseigenschaften linearer Differentialgleichungen 2. Ordnung, Mathematische Annalen, vol. 68 (1910), pp. 279–304, and vol. 71 (1911), pp. 214–232.

7. Hahn, Über Variationsprobleme mit variabeln Endpunkten, Monatshefte für Mathematik und Physik, vol. 22 (1911), pp. 127–136.

8. Bolza, Über den anormalen Fall beim Lagrangeschen und Mayerschen Problem mit gemischten Bedingungen und variablen Endpunkten, Mathematische Annalen, vol. 74 (1913), pp. 430–446.

[26] 11, Chapter VI, pp. 103–114.

9. The ideas of Hilbert concerning the relationships between boundary value problems and the calculus of variations have become accessible for the most part through the papers of his students, such as 4, 5, 6 above. For simpler cases see Lovitt, Linear Integral Equations, pp. 192–207.

10. Lecat, Calcul des variations, Encyclopédie des Sciences Mathématiques, II 31 (1913), pp. 156–158.

11. Hancock, Theory of Maxima and Minima, Ginn & Co., Boston (1917).

12. Bliss, The problem of Mayer with variable end-points, Transactions of the American Mathematical Society, vol. 19 (1918), pp. 305–314.

13. Bliss, A boundary value problem of the calculus of variations, Bulletin of the American Mathematical Society, vol. 32 (1926), pp. 317–331.

14. Cope, An analogue of Jacobi's condition for the problem of Mayer with variable end-points, Dissertation, The University of Chicago (1927).

15. Bliss, The problem of Lagrange in the calculus of variations, American Journal of Mathematics, vol. 52 (1930), pp. 673–744.

16. Morse and Myers, The problems of Lagrange and Mayer with variable end-points, Proceedings of the American Academy of Arts and Sciences, vol. 66 (1931), pp. 235–253.

17. Morse, Sufficient conditions in the problem of Lagrange with variable end conditions, American Journal of Mathematics, vol. 53 (1931), pp. 517–546.

18. Morse, Sufficient conditions in the problem of Lagrange with fixed end-points, Annals of Mathematics, vol. 32 (1931), pp. 567–577.

AN INEQUALITY FOR THE BESSEL FUNCTIONS OF THE FIRST KIND WITH IMAGINARY ARGUMENT.[1]

By T. H. Gronwall.

1. **Introduction.** In connection with a problem in wave mechanics, the question arose of finding upper and lower bounds of the expression

$$y = y_\nu(z) = z I_\nu'(z)/I_\nu(z), \tag{1}$$

where

$$I_\nu(z) = \sum_{m=0}^{\infty} \frac{(\frac{1}{2}z)^{\nu+2m}}{m!\,\Gamma(\nu+m+1)}, \tag{2}$$

and z and ν are both positive. Since z and ν may both be large, and their relative order of magnitude is not prescribed, the ordinary asymptotic methods are insufficient.

It is the main purpose of the present note to show that for $z>0$ and $\nu \geqq \frac{1}{2}$

$$z-\frac{1}{2}+\frac{\lambda}{2z}(1-e^{-2z}) < y_\nu(z) < z-\frac{1}{2}+\frac{\nu^2+\frac{3}{4}}{2z}(1-e^{-2z}), \tag{3}$$

where

$$\lambda = \begin{cases} \nu^2-\dfrac{1}{4}, & \dfrac{1}{2} \leqq \nu \leqq \dfrac{3}{2}, \\ \nu+\dfrac{1}{2}, & \nu \geqq \dfrac{3}{2}. \end{cases} \tag{4}$$

For many purposes, (3) may be replaced by the weaker inequality

$$z-\frac{1}{2} < y_\nu(z) < z-\frac{1}{2}+\frac{\nu^2+\frac{3}{4}}{2z}; \quad z>0, \quad \nu \geqq \frac{1}{2}. \tag{5}$$

In the special case where ν is fixed and z large, (3) is nearly as accurate as the asymptotic formula

$$y_\nu(z) = z-\frac{1}{2}+\frac{\nu^2-\frac{1}{4}}{2z}+O\left(\frac{1}{z^2}\right),$$

which is readily obtained from the familiar asymptotic expansions of $I_\nu(z)$ and $2I_\nu'(z) = I_{\nu-1}(z)+I_{\nu+1}(z)$.

2. **Proof of the inequality** (3). From the differential equation

$$z^2 I_\nu''(z)+z I_\nu'(z)-(z^2+\nu^2) I_\nu(z) = 0$$

[1] Received July 6, 1931.—Presented to the American Mathematical Society, Sept. 8, 1931.

and (1) it follows that

$$\frac{dy}{dz} = z + \frac{\nu^2 - y^2}{z}. \tag{6}$$

It is seen from (2) that y is finite and positive for $z > 0$, $\nu \geqq 0$, and for z sufficiently small

$$y = \nu + \frac{z^2}{2\nu + 2} + \cdots. \tag{7}$$

Writing

$$y = z - \frac{1}{2} + \frac{u}{z}, \tag{8}$$

so that for z small

$$u = \left(\nu + \frac{1}{2}\right) z + \cdots, \tag{9}$$

(6) becomes

$$\frac{du}{dz} = \nu^2 + \frac{3}{4} - 2u - \left(\frac{u}{z} - 1\right)^2. \tag{10}$$

Consequently

$$\frac{du}{dz} + 2u \leqq \nu^2 + \frac{3}{4}, \tag{11}$$

where, according to (10), the equality sign can hold at isolated points only. Multiplying this inequality by e^{2z}, integrating from 0 to z and observing that u vanishes for $z = 0$ by (9), we find for $z > 0$, $\nu \geqq 0$

$$u e^{2z} < \frac{1}{2}\left(\nu^2 + \frac{3}{4}\right)(e^{2z} - 1).$$

This, together with (8), gives the inequality to the right in (3).

To prove the inequality to the left in (3), we begin by showing that for $\nu \geqq \frac{1}{2}$

$$v = y - z \tag{12}$$

is $> -\frac{1}{2}$ and decreases monotonely as z increases.

From (12) and (6) we obtain

$$\frac{dv}{dz} = -1 - 2v + \frac{\nu^2 - v^2}{z}, \tag{13}$$

whence

$$\frac{d^2v}{dz^2} = -\frac{1 + 2v}{2} - \left(3 + \frac{2v}{z}\right)\frac{dv}{dz}. \tag{14}$$

For $z = 0$, $v = \nu$; letting z_1 be the smallest positive value of z for which $v = -\frac{1}{2}$, then obviously $dv/dz \leqq 0$ for $z = z_1$, and by (13) $\nu^2 \leqq \frac{1}{4}$. Since we assumed $\nu \geqq \frac{1}{2}$, we must have $\nu = \frac{1}{2}$, and in this

case, (13) and (14) show that all the derivatives of v vanish at $z = z_1$, so that v is constant $= -\frac{1}{2}$, which contradicts (7). By (12) and (7), v decreases as z increases for z sufficiently small; the first extreme, if any, must therefore be a minimum, and according to (14), this minimum must satisfy $1+2v \leqq 0$, whereas we have just shown that $v > -\frac{1}{2}$ for all $z > 0$.

Now $u/z = v+\frac{1}{2}$, and therefore when $\nu \geqq \frac{1}{2}$, u/z is positive and monotone decreasing for $z > 0$. Consequently, when $u/z \geqq 1$, we have

$$\left|\frac{u}{z}-1\right| = \frac{u}{z}-1 < \nu+\frac{1}{2}-1 = \nu-\frac{1}{2},$$

and when $u/z < 1$,

$$\left|\frac{u}{z}-1\right| = 1-\frac{u}{z} < 1.$$

Hence, by (10) and (4), we have for $z > 0$

$$\frac{du}{dz}+2u > \nu^2+\frac{3}{4}-\max.\left(\left(\nu-\frac{1}{2}\right)^2, 1\right) = \lambda, \tag{15}$$

and treating this inequality in the same way as (11), we obtain the inequality to the left in (3).

3. **Some other properties of $y_\nu(z)$.** We begin by showing that

$$\nu < y_\nu(z) < \sqrt{\nu^2+z^2}; \qquad \nu > 0,\ z > 0, \tag{16}$$

the differences $y-\nu$ and $\sqrt{\nu^2+z^2}-y$ both increasing monotonely with z. By (7), y increases with z for z sufficiently small, so that the first extreme of y, if any, is a maximum $> \nu$. But (6) shows that when $dy/dz = 0$, then

$$\frac{d^2y}{dz^2} = 1+\frac{y^2-\nu^2}{z^2},$$

so that $d^2y/dz^2 > 0$ for the maximum in question, which is impossible. Similarly, writing

$$w = \sqrt{\nu^2+z^2}-y,$$

we find from (7) that for small values of z

$$w = \left(\frac{1}{2\nu}-\frac{1}{2\nu+2}\right)z^2+\cdots$$

is positive and increasing, so that the first extreme, if any, is a maximum > 0. From (17) and (6) it is seen that when $dw/dz = 0$,

$$\frac{d^2w}{dz^2} = \frac{\nu^2}{(\nu^2+z^2)^{3/2}}+\frac{2\nu^2 w}{z^2(\nu^2+z^2)^{1/2}},$$

whence the same contradiction as above. We may note incidentally that the first part of (16), $y > \nu$, follows readily from the expansion of y in a continued fraction which is obtained from the recurrent formulas for $I_\nu(z)$.

We shall show finally that for $z \geqq 0$, $\nu \geqq 0$, y increases monotonely with ν, or

$$y_\mu(z) - y_\nu(z) > 0; \quad z \geqq 0, \quad \mu > \nu \geqq 0. \tag{17}$$

By (7), (17) holds for z sufficiently small; letting z_1 be the smallest positive value of z for which $y_\mu = y_\nu$, we must evidently have $y'_\mu(z_1) \leqq y'_\nu(z_1)$, whereas by (6) $y'_\mu(z_1) - y'_\nu(z_1) = (\mu^2 - \nu^2)/z_1 > 0$.

Since $I_{1/2}(z) = (2/\pi z)^{1/2} \operatorname{sh} z$, it follows from (17) that for $\nu \geqq \frac{1}{2}$ and $z > 0$

$$y_\nu(z) \geqq y_{1/2}(z) = z \coth z - \frac{1}{2} > z - \frac{1}{2},$$

which constitutes another proof of the first part of (5).

DEPARTMENT OF PHYSICS, COLUMBIA UNIVERSITY.

A SPECIAL CONFORMALLY EUCLIDEAN SPACE OF THREE DIMENSIONS OCCURRING IN WAVE MECHANICS.[1]

By T. H. Gronwall.

1. Introduction. Let

$$ds^2 = g_{ij}\,dx_i\,dx_j \tag{1}$$

define a Riemann space,[2] and let A and B be functions of the x_i. Consider the linear partial differential equation

$$\frac{\partial}{\partial x_i}\left(Ag^{ij}\frac{\partial \psi}{\partial x_j}\right)+AB\psi = 0 \tag{2}$$

which is self-adjoint and of elliptic type, since it is the Euler equation of the variation problem

$$\delta\int\cdots\int\left(g^{ij}\frac{\partial \psi}{\partial x_i}\frac{\partial \psi}{\partial x_j}-B\psi^2\right)A\,dx_1\cdots dx_n = 0. \tag{3}$$

The discussion of any boundary problem connected with (2) is greatly simplified when it is possible, by introducing new variables $\bar{x}_i$, to reduce the aggregate of the terms containing the second order derivatives to a multiple of the ordinary Laplacian:

$$\bar{A}\left(\frac{\partial^2 \psi}{\partial \bar{x}_1^2}+\frac{\partial^2 \psi}{\partial \bar{x}_2^2}+\cdots+\frac{\partial^2 \psi}{\partial \bar{x}_n^2}\right).$$

The necessary and sufficient condition for this is evidently that the space (1) be conformally euclidean:

$$g_{ij}\,dx_i\,dx_j = e^{-2\sigma}(d\bar{x}_1^2+d\bar{x}_2^2+\cdots+d\bar{x}_n^2). \tag{4}$$

It is the purpose of the present paper to apply this criterion to an equation which occurs in the wave mechanics of the helium atom.

We are dealing with a nucleus carrying Z charges ($Z = 2$ for the helium atom) and supposed at rest at the origin, and two electrons at distances r_1 and r_2 from it, their mutual distance being r_3. The corresponding wave equation is, using the "rational" units[3] of length and energy,

[1] Received August 3, 1931.—Presented to the American Mathematical Society, September 8, 1931.

[2] All variables and functions are real, and the quadratic form in (1) positive. We use subscripts instead of superscripts on the variables to simplify the typography.

[3] E. A. Hylleraas, Neue Berechnung der Energie des Heliums im Grundzustande, sowie des tiefsten Terms von Ortho-Helium. Zeitschrift für Physik **54** (1929) 347–366.

$$\nabla^2\psi+\left(\frac{E}{4}+\frac{1}{r_1}+\frac{1}{r_2}-\frac{1}{Zr_3}\right)\psi=0, \tag{5}$$

where the Laplacian is formed with respect to the six rectangular coördinates of the two electrons. Instead of these six coördinates, we take as variables r_1, r_2 and r_3 together with the three Euler angles which define the position in space of the triangle with vertices at the origin and at the two electrons. The dependence of the wave function ψ on these Euler angles is quite simple, being governed by the momentum equation,[4] which in our case is integrable in finite terms, so that the whole difficulty lies in investigating ψ as a function of the r's. The simplest case is here what the spectroscopists call the S states, where ψ is independent of the Euler angles, so that

$$\psi=\psi(r_1, r_2, r_3). \tag{6}$$

Substituting (6) in (5), an elementary calculation gives[5]

$$\begin{aligned}&\frac{\partial^2\psi}{\partial r_1^2}+\frac{\partial^2\psi}{\partial r_2^2}+2\frac{\partial^2\psi}{\partial r_3^2}+\frac{r_1^2-r_2^2+r_3^2}{r_1 r_3}\frac{\partial^2\psi}{\partial r_1\,\partial r_3}\\&+\frac{r_2^2-r_1^2+r_3^2}{r_2 r_3}\frac{\partial^2\psi}{\partial r_2\,\partial r_3}+\frac{2}{r_1}\frac{\partial\psi}{\partial r_1}+\frac{2}{r_2}\frac{\partial\psi}{\partial r_2}+\frac{4}{r_3}\frac{\partial\psi}{\partial r_3}\\&+\left(\frac{E}{4}+\frac{1}{r_1}+\frac{1}{r_2}-\frac{1}{Zr_3}\right)\psi=0.\end{aligned} \tag{7}$$

As noted by Hylleraas, this equation becomes self-adjoint upon multiplication by $r_1 r_2 r_3$, and the corresponding variation problem (3) is[6]

$$\begin{aligned}\delta\int\int\int\Big[&\left(\frac{\partial\psi}{\partial r_1}\right)^2+\left(\frac{\partial\psi}{\partial r_2}\right)^2+2\left(\frac{\partial\psi}{\partial r_3}\right)^2\\&+\frac{r_1^2-r_2^2+r_3^2}{r_1 r_3}\frac{\partial\psi}{\partial r_1}\frac{\partial\psi}{\partial r_3}+\frac{r_2^2-r_1^2+r_3^2}{r_2 r_3}\frac{\partial\psi}{\partial r_2}\frac{\partial\psi}{\partial r_3}\\&-\left(\frac{E}{4}+\frac{1}{r_1}+\frac{1}{r_2}-\frac{1}{Zr_3}\right)\psi^2\Big]r_1 r_2 r_3\,dr_1\,dr_2\,dr_3=0,\end{aligned} \tag{8}$$

the integration extending over all possible values of r_1, r_2 and r_3. Since they are the sides of a triangle, the sum of two is not less than the third, so that the region of integration is given by

$$r_1\geqq 0,\quad r_2\geqq 0,\quad |r_1-r_2|\leqq r_3\leqq r_1+r_2. \tag{9}$$

[4] This equation is a consequence of the invariance of (5), and its set of proper functions (Eigenfunktionen), under a rotation of the coördinate axes.

[5] Hylleraas, l. c. [3], 350, (5).

[6] L. c. [3], 351, (6).

We shall prove that a solution of the conformality condition (4) corresponding to (8) is

$$
\begin{aligned}
4x_1 &= r_1^2+r_2^2-r_3^2,\\
4x_2 &= r_1^2-r_2^2,\\
4x_3 &= \sqrt{(r_1+r_2+r_3)(r_1+r_2-r_3)(r_2+r_3-r_1)(r_3+r_1-r_2)};
\end{aligned}
\tag{10}
$$

in the expression for x_3, we recognize Heron's formula for the area of the triangle with sides r_1, r_2, r_3.

Transformed to the new variables, (8) becomes

$$
\begin{aligned}
\delta\int\int\int\Big[&\left(\frac{\partial\psi}{\partial x_1}\right)^2+\left(\frac{\partial\psi}{\partial x_2}\right)^2+\left(\frac{\partial\psi}{\partial x_3}\right)^2\\
&-\frac{1}{r}\left(\frac{E}{4}+\frac{1}{\sqrt{2(r+x_2)}}+\frac{1}{\sqrt{2(r-x_2)}}-\frac{1}{2Z\sqrt{r-x_1}}\right)\psi^2\Big]\\
&\times x_3\,dx_1\,dx_2\,dx_3=0,
\end{aligned}
\tag{11}
$$

where

$$
r=\sqrt{x_1^2+x_2^2+x_3^2}. \tag{12}
$$

The region of integration (9) takes the extremely simple form of the upper half space

$$
x_3\geqq 0, \tag{13}
$$

and the differential equation corresponding to (11) becomes upon division by x_3:

$$
\begin{aligned}
&\frac{\partial^2\psi}{\partial x_1^2}+\frac{\partial^2\psi}{\partial x_2^2}+\frac{\partial^2\psi}{\partial x_3^2}+\frac{1}{x_3}\frac{\partial\psi}{\partial x_3}\\
&+\frac{1}{r}\left(\frac{E}{4}+\frac{1}{\sqrt{2(r+x_2)}}+\frac{1}{\sqrt{2(r-x_2)}}-\frac{1}{2Z\sqrt{r-x_1}}\right)\psi=0.
\end{aligned}
\tag{15}
$$

On a former occasion,[7] the transformation (10) was but a means to an end; accordingly it was introduced without further explanation, and the transformation of (8) into (11) verified by a direct and very short calculation. At present, we shall follow the more arduous path which actually led to the discovery of (10).

2. **Preliminary transformations.** Our immediate object is to introduce new variables ξ_1, ξ_2, ξ_3 so as to get rid of the product terms in the quadratic form in (8). Inspection of (9) suggests as a reasonable choice of two of our new variables $\xi_1=(r_1+r_2)/r_3$ and $\xi_2=(r_1-r_2)/r_3$; the third, $\xi_3=\xi_3(r_1,r_2,r_3)$, is to be determined by the condition just mentioned. Transforming the bracket in (8) to the new variables, it is found to contain the following two product terms:

[7] T. H. Gronwall, The wave equation of the helium atom. I. Physical Review.

$$\left\{\frac{(r_2-r_1)\left[(r_1+r_2)^2-r_3^2\right]}{r_3^3}\left(\frac{1}{r_1}\frac{\partial\xi_3}{\partial r_1}-\frac{1}{r_2}\frac{\partial\xi_3}{\partial r_2}\right)\right.$$
$$\left.+\frac{(r_1+r_2)\left[r_3^2-(r_1-r_2)^2\right]}{r_1r_2r_3^2}\frac{\partial\xi_3}{\partial r_3}\right\}\frac{\partial\psi}{\partial\xi_1}\frac{\partial\psi}{\partial\xi_3}$$
$$+\left\{\frac{(r_1+r_2)\left[r_3^2-(r_1-r_2)^2\right]}{r_3^3}\left(\frac{1}{r_1}\frac{\partial\xi_3}{\partial r_1}-\frac{1}{r_2}\frac{\partial\xi_3}{\partial r_2}\right)\right.$$
$$\left.+\frac{(r_1-r_2)\left[(r_1+r_2)^2-r_3^2\right]}{r_1r_2r_3^2}\frac{\partial\xi_3}{\partial r_3}\right\}\frac{\partial\psi}{\partial\xi_2}\frac{\partial\psi}{\partial\xi_3}.$$

In order to make these vanish, we must have

$$\frac{1}{r_1}\frac{\partial\xi_3}{\partial r_1}-\frac{1}{r_2}\frac{\partial\xi_3}{\partial r_2}=0,\qquad \frac{\partial\xi_3}{\partial r_3}=0,$$

the solution of which is evidently $\xi_3=\varphi(r_1^2+r_2^2)$, where φ is an arbitrary function. We may thus choose as new variables

$$\xi_1=\frac{r_1+r_2}{r_3},\qquad \xi_2=\frac{r_1-r_2}{r_3},\qquad \xi_3=\sqrt{r_1^2+r_2^2},\tag{15}$$

and (8) is transformed into

$$\delta\int\int\int\left[(\xi_1^2-1)\left(\frac{\partial\psi}{\partial\xi_1}\right)^2+(1-\xi_2^2)\left(\frac{\partial\psi}{\partial\xi_2}\right)^2+\frac{(\xi_1^2-\xi_2^2)\xi_3^2}{(\xi_1^2+\xi_2^2)^2}\left(\frac{\partial\psi}{\partial\xi_3}\right)^2\right.$$
$$\left.-\frac{(\xi_1^2-\xi_2^2)\xi_3^2}{(\xi_1^2+\xi_2^2)^2}\left(\frac{E}{4}+\frac{1}{r_1}+\frac{1}{r_2}-\frac{1}{Zr_3}\right)\psi^2\right]\frac{\xi_3^2}{\xi_1^2+\xi_2^2}\,d\xi_1\,d\xi_2\,d\xi_3=0.\tag{16}$$

The further transformation

$$\xi_1=\operatorname{ch}u_1,\qquad \xi_2=\cos u_2,\qquad \xi_3=e^{u_3}\tag{17}$$

carries the quadratic form inside the bracket in (16) into

$$\left(\frac{\partial\psi}{\partial u_1}\right)^2+\left(\frac{\partial\psi}{\partial u_2}\right)^2+\frac{\operatorname{ch}^2u_1-\cos^2u_2}{(\operatorname{ch}^2u_1+\cos^2u_2)^2}\left(\frac{\partial\psi}{\partial u_3}\right)^2,\tag{18}$$

and taking this to be our $g^{ij}\frac{\partial\psi}{\partial u_i}\frac{\partial\psi}{\partial u_j}$, the ds^2 in (1) becomes

$$ds^2=du_1^2+du_2^2+\frac{(\operatorname{ch}^2u_1+\cos^2u_2)^2}{\operatorname{ch}^2u_1-\cos^2u_2}\,du_3^2.\tag{19}$$

3. The Ricci tensor R_{ij} and proof that (19) is conformally euclidean. In the special case when ds^2 has the form

$$ds^2=g_{11}\,du_1^2+g_{22}\,du_2^2+g_{33}\,du_3^2,\tag{20}$$

the Christoffel symbols of the second kind are given by E(15.7),[8] and the equations of definition E(8.14) give the following values of the components of the Ricci tensor, where i, j, k denote a permutation of $1, 2, 3$:

$$
\begin{aligned}
R_{ii} = {} & \frac{1}{2g_{jj}}\left(\frac{\partial^2 g_{jj}}{\partial u_i^2}+\frac{\partial^2 g_{ii}}{\partial u_j^2}\right)+\frac{1}{2g_{kk}}\left(\frac{\partial^2 g_{kk}}{\partial u_i^2}+\frac{\partial^2 g_{ii}}{\partial u_k^2}\right) \\
& -\frac{1}{4g_{jj}^2}\left[\left(\frac{\partial g_{jj}}{\partial u_i}\right)^2+\frac{\partial g_{ii}}{\partial u_j}\,\frac{\partial g_{jj}}{\partial u_j}\right] \\
& -\frac{1}{4g_{kk}^2}\left[\left(\frac{\partial g_{kk}}{\partial u_i}\right)^2+\frac{\partial g_{ii}}{\partial u_k}\,\frac{\partial g_{kk}}{\partial u_k}\right] \\
& -\frac{1}{4g_{ii}\,g_{jj}}\left[\frac{\partial g_{ii}}{\partial u_i}\,\frac{\partial g_{jj}}{\partial u_i}+\left(\frac{\partial g_{ii}}{\partial u_j}\right)^2\right] \\
& -\frac{1}{4g_{ii}\,g_{kk}}\left[\frac{\partial g_{ii}}{\partial u_i}\,\frac{\partial g_{kk}}{\partial u_i}+\left(\frac{\partial g_{ii}}{\partial u_k}\right)^2\right] \\
& +\frac{1}{4g_{jj}\,g_{kk}}\left[\frac{\partial g_{ii}}{\partial u_j}\,\frac{\partial g_{kk}}{\partial u_j}+\frac{\partial g_{ii}}{\partial u_k}\,\frac{\partial g_{jj}}{\partial u_k}\right], \\
R_{ij} = {} & \frac{1}{2}\,\frac{\partial^2 \log g_{kk}}{\partial u_i\,\partial u_j}+\frac{1}{4}\,\frac{\partial \log g_{kk}}{\partial u_i}\,\frac{\partial \log g_{kk}}{\partial u_j} \\
& -\frac{1}{4}\,\frac{\partial \log g_{ii}}{\partial u_j}\,\frac{\partial \log g_{kk}}{\partial u_i}-\frac{1}{4}\,\frac{\partial \log g_{jj}}{\partial u_i}\,\frac{\partial \log g_{kk}}{\partial u_j}.
\end{aligned}
$$

We specialize (20) further by making

$$
g_{11} = g_{22} = 1, \qquad \frac{\partial g_{33}}{\partial u_3} = 0; \tag{21}
$$

since we now have

$$
g = g_{33}, \tag{22}
$$

we may drop the subscripts on g_{33}, and the preceding formulas give the following values of the Ricci tensor components, where those not written out vanish:

$$
\begin{aligned}
R_{11} &= \frac{1}{2}\,\frac{\partial^2 \log g}{\partial u_1^2}+\frac{1}{4}\left(\frac{\partial \log g}{\partial u_1}\right)^2, \\
R_{22} &= \frac{1}{2}\,\frac{\partial^2 \log g}{\partial u_2^2}+\frac{1}{4}\left(\frac{\partial \log g}{\partial u_2}\right)^2, \\
R_{12} &= R_{21} = \frac{1}{2}\,\frac{\partial^2 \log g}{\partial u_1\,\partial u_2}+\frac{1}{4}\,\frac{\partial \log g}{\partial u_1}\,\frac{\partial \log g}{\partial u_2}, \\
R_{33} &= \frac{1}{2}\,\frac{\partial^2 g}{\partial u_1^2}+\frac{1}{2}\,\frac{\partial^2 g}{\partial u_2^2}-\frac{1}{4g}\left(\frac{\partial g}{\partial u_1}\right)^2-\frac{1}{4g}\left(\frac{\partial g}{\partial u_2}\right)^2.
\end{aligned} \tag{23}
$$

[8] The letter E followed by the number of an equation or a page indicates a reference to L. P. Eisenhart, Riemannian Geometry, Princeton University Press, 1926. We use Eisenhart's notations throughout except that x^i is replaced by u_i.

It is convenient also to write these equations in the following form:

$$(24)\qquad \begin{aligned} R_{11} &= \frac{1}{\sqrt{g}}\frac{\partial^2\sqrt{g}}{\partial u_1^2}, \quad R_{12} = R_{21} = \frac{1}{\sqrt{g}}\frac{\partial^2\sqrt{g}}{\partial u_1\,\partial u_2}, \\ R_{22} &= \frac{1}{\sqrt{g}}\frac{\partial^2\sqrt{g}}{\partial u_2^2}, \quad R_{33} = g(R_{11}+R_{22}). \end{aligned}$$

A necessary and sufficient condition that a space in three dimensions can be mapped conformally on a flat space is given in E (theorem p. 92) and consists in the vanishing of the tensor R_{ijk} defined by E (28.14). In the special case (20), these equations become, denoting again by i, j, k a permutation of 1, 2, 3:

$$R_{iii} = 0, \qquad R_{ijj} = 0,$$

$$R_{ijk} = \frac{\partial R_{ij}}{\partial u_k} - \frac{\partial R_{ik}}{\partial u_j},$$

$$\begin{aligned} R_{iij} &= -R_{iji} \\ &= \frac{\partial R_{ii}}{\partial x_j} - \frac{\partial R_{ij}}{\partial x_i} - R_{ii}\left\{\begin{matrix} i \\ i\,j \end{matrix}\right\} + R_{jj}\left\{\begin{matrix} j \\ i\,i \end{matrix}\right\} - \frac{1}{4}g_{ii}\frac{\partial R}{\partial x_j}, \end{aligned}$$

where $R = g^{11}R_{11} + g^{22}R_{22} + g^{33}R_{33}$. Specializing to (21), we have by (24)

$$(25)\qquad R = 2(R_{11}+R_{22}),$$

and the conditions for the vanishing of our tensor reduce to the two equations

$$R_{112} = \frac{1}{2}\left(\frac{\partial R_{11}}{\partial u_2} - \frac{\partial R_{22}}{\partial u_1}\right) - \frac{\partial R_{12}}{\partial u_1} = 0,$$

$$R_{221} = \frac{1}{2}\left(\frac{\partial R_{22}}{\partial u_1} - \frac{\partial R_{11}}{\partial u_2}\right) - \frac{\partial R_{21}}{\partial u_2} = 0.$$

By means of (23), these may be written

$$\frac{\partial}{\partial u_i}\left[g\left(\frac{\partial^2 \log g}{\partial u_1^2} + \frac{\partial^2 \log g}{\partial u_2^2}\right)\right] = 0, \qquad (i = 1, 2),$$

and since g is independent of u_3, these equations are equivalent to

$$(26)\qquad \frac{\partial^2 \log g}{\partial u_1^2} + \frac{\partial^2 \log g}{\partial u_2^2} = \frac{c}{g},$$

where c is a constant. Thus (26) is the condition that (19) be conformal, or, rewriting (4) in the new notation, that

$$(27)\qquad ds^2 = du_1^2 + du_2^2 + g\,du_3^2 = e^{-2\sigma}(dx_1^2 + dx_2^2 + dx_3^2),$$

where

$$g = \frac{(\mathrm{ch}^2 u_1 + \cos^2 u_2)^2}{\mathrm{ch}^2 u_1 - \cos^2 u_2}. \tag{28}$$

Now (28) satisfies (26) with $c = 8$ as is seen by direct calculation.

It is convenient to rewrite equation (26) in the form

$$\sqrt{g}\left(\frac{\partial^2 \sqrt{g}}{\partial u_1^2} + \frac{\partial^2 \sqrt{g}}{\partial u_2^2}\right) - \left(\frac{\partial \sqrt{g}}{\partial u_1}\right)^2 - \left(\frac{\partial \sqrt{g}}{\partial u_2}\right)^2 = \frac{1}{2} g \left(\frac{\partial^2 \log g}{\partial u_1^2} + \frac{\partial^2 \log g}{\partial u_2^2}\right) = 4. \tag{29}$$

For future reference, we calculate the following expressions from (23):

$$R_{11} = \frac{3 - 4\,\mathrm{ch}^2 u_1}{\mathrm{ch}^2 u_1 - \cos^2 u_2} + \frac{3\,\mathrm{ch}^2 u_1 (\mathrm{ch}^2 u_1 - 1)}{(\mathrm{ch}^2 u_1 - \cos^2 u_2)^2} + \frac{2\,\mathrm{ch}^2 u_1}{\mathrm{ch}^2 u_1 + \cos^2 u_2},$$

$$R_{12} = \frac{3\,\mathrm{ch}\, u_1\, \mathrm{sh}\, u_1 \cos u_2 \sin u_2}{(\mathrm{ch}^2 u_1 - \cos^2 u_2)^2}, \tag{30}$$

$$R_{22} = \frac{3 - 4\cos^2 u_2}{\mathrm{ch}^2 u_1 - \cos^2 u_2} + \frac{3\cos^2 u_2 (1 - \cos^2 u_2)}{(\mathrm{ch}^2 u_1 - \cos^2 u_2)^2} - \frac{2\cos^2 u_2}{\mathrm{ch}^2 u_1 + \cos^2 u_2}.$$

We may note in passing that (27) is not a ds^2 of constant curvature; one of the conditions E (27.1) for constant curvature K_0 is

$$R_{1332} = K_0 (g_{13}\, g_{32} - g_{12}\, g_{33}),$$

and the expression to the right vanishes, whereas it is readily seen from E (15.8) that

$$R_{1332} = \sqrt{g}\, \frac{\partial^2 \sqrt{g}}{\partial u_1\, \partial u_2} \neq 0.$$

4. **The six differential equations which determine σ in (27).** They are E (28.18), and replacing the covariant derivatives by their explicit expressions, these equations become

$$\frac{\partial^2 \sigma}{\partial u_i\, \partial u_j} - \frac{\partial \sigma}{\partial u_i}\frac{\partial \sigma}{\partial u_j} - \left\{\begin{matrix} h \\ i\,j \end{matrix}\right\} \frac{\partial \sigma}{\partial u_h} - \left(\frac{1}{2} R g_{ij} - R_{ij}\right) + \frac{1}{2} g_{ij}\, g^{hk} \frac{\partial \sigma}{\partial u_h}\frac{\partial \sigma}{\partial u_k} = 0.$$

Upon multiplication by $-e^{-\sigma}$, these equations may be written

$$\frac{\partial^2 e^{-\sigma}}{\partial u_i\, \partial u_j} - \left\{\begin{matrix} h \\ i\,j \end{matrix}\right\} \frac{\partial e^{-\sigma}}{\partial u_h} + \left(\frac{1}{2} R g_{ij} - R_{ij}\right) e^{-\sigma} - \frac{1}{2} e^{\sigma} g_{ij}\, g^{hk} \frac{\partial e^{-\sigma}}{\partial u_h}\frac{\partial e^{-\sigma}}{\partial u_k} = 0.$$

Specializing to (27), and using (24) and (25), we find

$$\frac{\partial}{\partial u_1}\left(\frac{1}{\sqrt{g}}\frac{\partial e^{-\sigma}}{\partial u_3}\right)=0,\quad \frac{\partial}{\partial u_2}\left(\frac{1}{\sqrt{g}}\frac{\partial e^{-\sigma}}{\partial u_3}\right)=0,$$

$$\frac{\partial^2 e^{-\sigma}}{\partial u_1\,\partial u_2}-R_{12}e^{-\sigma}=0,$$

$$\begin{aligned}
&\frac{\partial^2 e^{-\sigma}}{\partial u_1^2}+\frac{1}{2}(R_{22}-R_{11})e^{-\sigma}\\
&\qquad-\frac{1}{2}e^{\sigma}\left[\left(\frac{\partial e^{-\sigma}}{\partial u_1}\right)^2+\left(\frac{\partial e^{-\sigma}}{\partial u_2}\right)^2+\frac{1}{g}\left(\frac{\partial e^{-\sigma}}{\partial u_3}\right)^2\right]=0,\\
&\frac{\partial^2 e^{-\sigma}}{\partial u_2^2}+\frac{1}{2}(R_{11}-R_{22})e^{-\sigma}\\
&\qquad-\frac{1}{2}e^{\sigma}\left[\left(\frac{\partial e^{-\sigma}}{\partial u_1}\right)^2+\left(\frac{\partial e^{-\sigma}}{\partial u_2}\right)^2+\frac{1}{g}\left(\frac{\partial e^{-\sigma}}{\partial u_3}\right)^2\right]=0,\\
&\frac{\partial^2 e^{-\sigma}}{\partial u_3^2}+\frac{1}{2}\frac{\partial g}{\partial u_1}\frac{\partial e^{-\sigma}}{\partial u_1}+\frac{1}{2}\frac{\partial g}{\partial u_2}\frac{\partial e^{-\sigma}}{\partial u_2}-\frac{1}{2}g(R_{11}+R_{22})e^{-\sigma}\\
&\qquad-\frac{1}{2}e^{\sigma}\left[g\left(\frac{\partial e^{-\sigma}}{\partial u_1}\right)^2+g\left(\frac{\partial e^{-\sigma}}{\partial u_2}\right)^2+\left(\frac{\partial e^{-\sigma}}{\partial u_3}\right)^2\right]=0.
\end{aligned}\tag{31}$$

The general solution of the first two of these equations is evidently

$$e^{-\sigma}=F(u_3)\sqrt{g}+f(u_1,u_2).\tag{32}$$

Substituting this in the third of (31) and using (24), we obtain

$$\frac{\partial^2 f}{\partial u_1\,\partial u_2}-R_{12}f=0.\tag{33}$$

Subtraction of the fifth from the fourth in (31) gives

$$\frac{\partial^2 e^{-\sigma}}{\partial u_1^2}-\frac{\partial^2 e^{-\sigma}}{\partial u_2^2}+(R_{22}-R_{11})e^{-\sigma}=0,$$

which reduces to

$$\frac{\partial^2 f}{\partial u_1^2}-\frac{\partial^2 f}{\partial u_2^2}+(R_{22}-R_{11})f=0\tag{34}$$

by means of (32) and (24). Now add the fourth and fifth in (31), and multiply by $e^{-\sigma}$:

$$e^{-\sigma}\left(\frac{\partial^2 e^{-\sigma}}{\partial u_1^2}+\frac{\partial^2 e^{-\sigma}}{\partial u_2^2}\right)-\left[\left(\frac{\partial e^{-\sigma}}{\partial u_1}\right)^2+\left(\frac{\partial e^{-\sigma}}{\partial u_2}\right)^2+\frac{1}{g}\left(\frac{\partial e^{-\sigma}}{\partial u_3}\right)^2\right]=0.$$

Substitution of (32) gives

$$(F\sqrt{g}+f)\left[F\left(\frac{\partial^2\sqrt{g}}{\partial u_1^2}+\frac{\partial^2\sqrt{g}}{\partial u_2^2}\right)+\frac{\partial^2 f}{\partial u_1^2}+\frac{\partial^2 f}{\partial u_2^2}\right]$$
$$-F^2\left[\left(\frac{\partial\sqrt{g}}{\partial u_1}\right)^2+\left(\frac{\partial\sqrt{g}}{\partial u_2}\right)^2\right]-2F\left[\frac{\partial\sqrt{g}}{\partial u_1}\frac{\partial f}{\partial u_1}+\frac{\partial\sqrt{g}}{\partial u_2}\frac{\partial f}{\partial u_2}\right]$$
$$-\left(\frac{\partial f}{\partial u_1}\right)^2-\left(\frac{\partial f}{\partial u_2}\right)^2-\left(\frac{dF}{du_3}\right)^2=0;$$

multiplying out, and reducing the coefficient of F^2 by (29) and that of F by (24), we find

$$4F^2+F\sqrt{g}\left[\frac{\partial^2 f}{\partial u_1^2}+\frac{\partial^2 f}{\partial u_2^2}-\frac{\partial\log g}{\partial u_1}\frac{\partial f}{\partial u_1}-\frac{\partial\log g}{\partial u_2}\frac{\partial f}{\partial u_2}+(R_{11}+R_{22})f\right]$$
$$+f\left(\frac{\partial^2 f}{\partial u_1^2}+\frac{\partial^2 f}{\partial u_2^2}\right)-\left(\frac{\partial f}{\partial u_1}\right)^2-\left(\frac{\partial f}{\partial u_2}\right)^2-\left(\frac{dF}{du_3}\right)^2=0. \tag{35}$$

Giving constant values to u_1 and u_2 this becomes

$$\left(\frac{dF}{du_3}\right)^2=4F^2+\alpha F+\beta, \tag{36}$$

where α and β are constants, and since adding a constant to F does not change the form of (32), we may assume $\alpha=0$, so that (36) gives

$$\frac{d^2F}{du_3^2}=4F,$$

the general solution of which is

$$F=ae^{2u_3}+be^{-2u_3}, \tag{37}$$

and (36) takes the form

$$\left(\frac{dF}{du_3}\right)^2=4F^2-16ab. \tag{38}$$

By means of (38), (35) splits up into two equations:

$$\frac{\partial^2 f}{\partial u_1^2}+\frac{\partial^2 f}{\partial u_2^2}-\frac{\partial\log g}{\partial u_1}\frac{\partial f}{\partial u_1}-\frac{\partial\log g}{\partial u_2}\frac{\partial f}{\partial u_2}+(R_{11}+R_{22})f=0, \tag{39}$$

$$f\left(\frac{\partial^2 f}{\partial u_1^2}+\frac{\partial^2 f}{\partial u_2^2}\right)-\left(\frac{\partial f}{\partial u_1}\right)^2-\left(\frac{\partial f}{\partial u_2}\right)^2+16ab=0. \tag{40}$$

Finally, multiply the last of (31) by $e^{-\sigma}$, substitute (32) and use successively (38), (29), (40), (39) and (24); the result is an identity.

There now remain to be solved equations (33), (34), (39) and (40); adding and subtracting (34) and (39), we obtain the system of linear equations

$$\frac{\partial^2 f}{\partial u_1^2} = \frac{1}{2}\frac{\partial \log g}{\partial u_1}\frac{\partial f}{\partial u_1} + \frac{1}{2}\frac{\partial \log g}{\partial u_2}\frac{\partial f}{\partial u_2} - R_{22} f,$$

$$\frac{\partial^2 f}{\partial u_1 \partial u_2} = R_{12} f, \tag{41}$$

$$\frac{\partial^2 f}{\partial u_2^2} = \frac{1}{2}\frac{\partial \log g}{\partial u_1}\frac{\partial f}{\partial u_1} + \frac{1}{2}\frac{\partial \log g}{\partial u_2} - R_{11} f.$$

Multiplying (39) by f, and subtracting (40), we have

$$\left(\frac{\partial f}{\partial u_1}\right)^2 + \left(\frac{\partial f}{\partial u_2}\right)^2 - \frac{\partial \log g}{\partial u_1} f \frac{\partial f}{\partial u_1} - \frac{\partial \log g}{\partial u_2} f \frac{\partial f}{\partial u_2} + (R_{11} + R_{22}) f^2 - 16\,ab = 0,$$

which, by the aid of (24) and (29), may be written

$$\left(\frac{\partial f}{\partial u_1} - \frac{f}{\sqrt{g}}\frac{\partial \sqrt{g}}{\partial u_1}\right)^2 + \left(\frac{\partial f}{\partial u_2} - \frac{f}{\sqrt{g}}\frac{\partial \sqrt{g}}{\partial u_2}\right)^2 + \frac{4f^2}{g} - 16\,ab = 0. \tag{42}$$

In the discussion of (41), it is convenient to change variables so as to make the coefficients rational functions: we therefore return to the variables (17), or preferably their squares, writing

$$t_1 = \operatorname{ch}^2 u_1 = \xi_1^2, \qquad t_2 = \cos^2 u_2 = \xi_2^2. \tag{43}$$

Using (30), this transforms (41) and (42) into

$$4t_1(t_1-1)\frac{\partial^2 f}{\partial t_1^2} + \left[2(2t_1-1) + \frac{2t_1(t_1-1)}{t_1-t_2} - \frac{4t_1(t_1-1)}{t_1+t_2}\right]\frac{\partial f}{\partial t_1} - \left[\frac{2t_2(1-t_2)}{t_1-t_2} + \frac{4t_2(1-t_2)}{t_1+t_2}\right]\frac{\partial f}{\partial t_2} + \left[\frac{3-4t_2}{t_1-t_2} + \frac{3t_2(1-t_2)}{(t_1-t_2)^2} - \frac{2t_2}{t_1+t_2}\right] f = 0,$$

$$4\frac{\partial^2 f}{\partial t_1 \partial t_2} + \frac{3}{(t_1-t_2)^2} f = 0, \tag{41'}$$

$$4t_2(1-t_2)\frac{\partial^2 f}{\partial t_2^2} + \left[\frac{2t_1(t_1-1)}{t_1-t_2} - \frac{4t_1(t_1-1)}{t_1+t_2}\right]\frac{\partial f}{\partial t_1} - \left[2(2t_2-1) + \frac{2t_2(1-t_2)}{t_1-t_2} + \frac{4t_2(1-t_2)}{t_1+t_2}\right]\frac{\partial f}{\partial t_2} + \left[\frac{3-4t_1}{t_1-t_2} + \frac{3t_1(t_1-1)}{(t_1-t_2)^2} + \frac{2t_1}{t_1+t_2}\right] f = 0;$$

$$t_1(t_1-1)\left[\frac{\partial f}{\partial t_1} + \left(\frac{1}{2}\frac{1}{t_1-t_2} - \frac{1}{t_1+t_2}\right) f\right]^2 + t_2(1-t_2)\left[\frac{\partial f}{\partial t_2} - \left(\frac{1}{2}\frac{1}{t_1-t_2} + \frac{1}{t_1+t_2}\right) f\right]^2 + \frac{t_1-t_2}{(t_1+t_2)^2} f^2 - 4\,ab = 0. \tag{42'}$$

One way of dealing with (41′) would be to solve the first equation for $\partial f/\partial t_2$ and substitute in the second, thus obtaining a linear differential equation of the third order in the variable t_1. After determining a fundamental system of solutions, the general solution is a linear combination of these with coefficients dependent on t_2, and these coefficients are determined by the last two of (41′). Instead of performing these rather tedious operations, we follow a suggestion from the general theory of completely integrable linear systems[9] and substitute an expansion in powers of $t_1 - t_2$ with coefficients dependent on t_2

$$f = \sum_{\nu=0}^{\infty} a_\nu(t_2)\,(t_1 - t_2)^{\varrho+\nu}, \qquad a_0(t_2) \neq 0,$$

in the first of (41′). The lowest power of $t_1 - t_2$ occurring is the $(\varrho-2)$th, and setting its coefficient equal to zero, we find

$$t_2(t_2-1)\,a_0(t_2)\,[4\varrho(\varrho-1)+2\varrho-2\varrho-3] = 0,$$

or

$$4\varrho(\varrho-1)-3 = 0, \quad \text{roots } \varrho = -\frac{1}{2} \text{ and } \frac{3}{2}.$$

This is the characteristic equation at the singularity $t_1 = t_2$, and the root $-\frac{1}{2}$ immediately suggests the substitution

$$f = (t_1 - t_2)^{-\frac{1}{2}}\,\varphi, \tag{44}$$

the result of which is[10]

$$2t_1(t_1-1)\frac{\partial^2\varphi}{\partial t_1^2} + \left[2t_1 - 1 - \frac{t_1(t_1-1)}{t_1-t_2} - \frac{2t_1(t_1-1)}{t_1+t_2}\right]\frac{\partial\varphi}{\partial t_1}$$
$$-\left[\frac{t_2(1-t_2)}{t_1-t_2} + \frac{2t_2(1-t_2)}{t_1+t_2}\right]\frac{\partial\varphi}{\partial t_2} + \frac{t_1-t_2}{t_1+t_2}\varphi = 0,$$

$$2\frac{\partial^2\varphi}{\partial t_1\,\partial t_2} + \frac{1}{t_1-t_2}\left(\frac{\partial\varphi}{\partial t_1} - \frac{\partial\varphi}{\partial t_2}\right) = 0, \tag{41″}$$

$$2t_2(1-t_2)\frac{\partial^2\varphi}{\partial t_2^2} + \left[\frac{t_1(t_1-1)}{t_1-t_2} - \frac{2t_1(t_1-1)}{t_1+t_2}\right]\frac{\partial\varphi}{\partial t_1}$$
$$+\left[1-2t_2+\frac{t_2(1-t_2)}{t_1-t_2} - \frac{2t_2(1-t_2)}{t_1+t_2}\right]\frac{\partial\varphi}{\partial t_2} + \frac{t_1-t_2}{t_1+t_2}\varphi = 0;$$

[9] For references to this theory, of which no further use is made in the text, see footnotes 122 and 128 of the article II B 5 (E.Hilb), Encycl. d. Math. Wiss. vol. 2, part 2, pp. 502-503.

[10] The work is simplified by the remark that permuting t_1 and t_2 merely permutes the first and third of (41′), so that we can substitute in the first two only and obtain the third equation (41″) by symmetry from the first.

$$(42'') \qquad \frac{t_1(t_1-1)}{t_1-t_2}\left(\frac{\partial\varphi}{\partial t_1}-\frac{\varphi}{t_1+t_2}\right)^2+\frac{t_2(1-t_2)}{t_1-t_2}\left(\frac{\partial\varphi}{\partial t_2}-\frac{\varphi}{t_1+t_2}\right)^2 + \left(\frac{\varphi}{t_1+t_2}\right)^2-4ab=0.$$

Previously, we formed the characteristic equation of (41′) at $t_1=t_2$; we now do the same with (41″) at $t_1=0, 1, \infty$ and find

$$\text{at } t_1=0, \quad -2\varrho(\varrho-1)-\varrho=0, \quad \text{roots } \varrho=0 \text{ and } \frac{1}{2},$$

$$\text{„ } t_1=1, \quad 2\varrho(\varrho-1)+\varrho=0, \quad \text{„ } \varrho=0 \text{ „ } \frac{1}{2},$$

$$\text{„ } t_1=\infty, \quad 2\varrho(\varrho-1)-\varrho+1=0, \quad \text{„ } \varrho=1 \text{ „ } \frac{1}{2}.$$

By symmetry,[10] the characteristic equations at $t_2=0, 1, \infty$ are the same as above. Now let us try to find a polynomial solution of (41″); since $\varrho=1$ at $t_1=\infty$ and at $t_2=\infty$, it must be linear in t_1 and in t_2. Substituting $\varphi=\alpha+\beta t_1+\gamma t_2+\delta t_1 t_2$ in the second of (41″), we find $\beta=\gamma$ and $\delta=0$, the first gives $\alpha=-2\beta$, and substitution in the third is superfluous on account of the symmetry. Hence

$$\varphi_1=t_1+t_2-2$$

is a solution of (41″). Inspection of the roots ϱ shows that $t_1^{\frac{1}{2}}t_2^{\frac{1}{2}}$ belongs to a possible set of exponents at $t_1, t_2=0, 1, \infty$, and substitution in the first two of (41″) shows that it is a solution:

$$\varphi_2=t_1^{\frac{1}{2}}t_2^{\frac{1}{2}}.$$

A similar argument yields a third solution

$$\varphi_3=(t_1-1)^{\frac{1}{2}}(1-t_2)^{\frac{1}{2}}.$$

Having thus exhibited three linearly independent solutions of (41″), we know that this system is completely integrable, the general solution being

$$(45) \qquad \varphi=c_1(t_1+t_2-2)+2c_2t_1^{\frac{1}{2}}t_2^{\frac{1}{2}}+2c_3(t_1-1)^{\frac{1}{2}}(1-t_2)^{\frac{1}{2}}$$

with constant c's. Finally, substitution of (45) in (42″) reduces the latter equation to

$$(46) \qquad c_1^2+c_2^2+c_3^2-4ab=0.$$

Collecting the information contained in (32), (37), (43), (44) and (45), we find that the general solution of (31) is

$$e^{-\sigma} = (\mathrm{ch}^2 u_1 - \cos^2 u_2)^{-\frac{1}{2}} [c_1 (\mathrm{ch}^2 u_1 + \cos^2 u_2 - 2) + 2 c_2 \,\mathrm{ch}\, u_1 \cos u_2 + 2 c_3 \,\mathrm{sh}\, u_1 \sin u_2 + (a e^{2u_3} + b e^{-2u_3}) (\mathrm{ch}^2 u_1 + \cos^2 u_2)], \tag{47}$$

with the relation (46) between the constants.

5. **Determination of the coördinates x_1, x_2, x_3 of the flat space** in (27). The differential equations for the x_i are readily obtained from $E(7.14)$ in the form

$$\frac{\partial^2 x_i}{\partial u_j \partial u_k} = \left\{ \begin{matrix} h \\ j k \end{matrix} \right\}' \frac{\partial x_i}{\partial u_h}, \tag{48}$$

where the prime on the Christoffel symbols indicates that they are formed with $g'_{ij} = e^{2\sigma} g_{ij}$. These equations are completely integrable for any values of the constants in (47), subject to (46); in fact, the integrability conditions of (48) are precisely (31), as is seen from the way they originated in $E(25)$.

This remark, and the familiar fact that an inversion maps a flat space conformally upon itself, enables us to dispense with the formal integration of (48), the solution being obtained instead by a simple algebraic calculation. Specialize (46) and (47) to $c_1 = c_2 = c_3 = 0$, $a = 0$, $b = 2$, and write

$$e^{-\sigma_0} = 2 e^{-2u_3} (\mathrm{ch}^2 u_1 + \cos^2 u_2) (\mathrm{ch}^2 u_1 - \cos^2 u_2)^{-\frac{1}{2}}, \tag{49}$$

$$dx_1^2 + dx_2^2 + dx_3^2 = e^{2\sigma_0} (du_1^2 + du_2^2 + g\, du_3^2); \tag{50}$$

since (48) is completely integrable in the special case (49), x_1, x_2 and x_3 are determined by (50) as functions of u_1, u_2 and u_3, and the determination is complete within a transformation leaving the left side of (50) invariant, i. e. a translation, rotation and reflexion of the flat space. Now we perform an inversion with respect to the sphere $\sum_h (x_h - a_h)^2 = R^2$, where a_1, a_2, a_3 and R are arbitrary parameters:

$$x'_i - a_i = \frac{R^2 (x_i - a_i)}{\sum_h (x_h - a_h)^2}. \tag{51}$$

By (51) and (50)

$$\begin{aligned} dx_1'^2 + dx_2'^2 + dx_3'^2 &= \frac{R^4}{\left[\sum_h (x_h - a_h)^2\right]^2} (dx_1^2 + dx_2^2 + dx_3^2) \\ &= \frac{R^4 e^{2\sigma_0}}{\left[\sum_h (x_h - a_h)^2\right]^2} (du_1^2 + du_2^2 + g\, du_3^2), \end{aligned} \tag{52}$$

and since this maps the given space conformally on the flat space x'_1, x'_2, x'_3, the expression to the right must be equal to

$$e^{2\sigma}(du_1^2+du_2^2+g\,du_3^2), \tag{53}$$

where now the constants in (47) depend on the parameters a_1, a_2, a_3 and R of the inversion. Comparison of (52) and (53) gives

$$R^2 e^{-\sigma} = e^{-\sigma_0}\sum_h (x_h-a_h)^2,$$

whence, introducing (47) and (49),

$$\begin{aligned} &R^2[c_1(\mathrm{ch}^2 u_1+\cos^2 u_2-2)+2c_2\,\mathrm{ch}\,u_1\cos u_2+2c_3\,\mathrm{sh}\,u_1\sin u_2 \\ &\quad+(a e^{2u_3}+b e^{-2u_3})(\mathrm{ch}^2 u_1+\cos^2 u_2)] \\ &\quad= 2e^{-2u_3}(\mathrm{ch}^2 u_1+\cos^2 u_2)\sum_h (x_h-a_h)^2. \end{aligned} \tag{54}$$

Since the x's are independent of the parameters, and the four functions of the u's in the bracket are linearly independent, (54) gives

$$\begin{aligned} &\frac{\partial R^2 c_k}{\partial R}=0,\quad (k=1,2,3),\quad \frac{\partial R^2 a}{\partial R}=0,\quad \frac{\partial R^2 b}{\partial R}=0,\\ &\frac{\partial^2 R^2 c_k}{\partial a_i\,\partial a_j}=0,\quad \frac{\partial^2 R^2 a}{\partial a_i\,\partial a_j}=0,\quad (i,j,k=1,2,3),\\ &\frac{\partial^2 R^2 b}{\partial a_i^2}=4,\quad \frac{\partial^2 R^2 b}{\partial a_i\,\partial a_j}=0,\quad (i\neq j), \end{aligned}$$

and integrating these equations,

$$\begin{aligned} R^2 c_k &= \sum_h \gamma_{hk} a_h+\gamma_k, && (k=1,2,3),\\ R^2 a &= \sum_h \alpha_h a_h+\alpha, \\ R^2 b &= 2\Big[\sum_h (a_h^2+\beta_h a_h)+\beta\Big], \end{aligned} \tag{55}$$

where all the Greek letters are real constants. By a suitable translation of our flat space, we may make $\beta_1=\beta_2=\beta_3=0$, and substituting (55) in (46), the terms of third degree in a_1, a_2, a_3 show that $\alpha_1=\alpha_2=\alpha_3=0$. With these simplifications, (46) becomes

$$\sum_{hik}\gamma_{hk}\gamma_{ik}a_h a_i-8\alpha\sum_h a_h^2+2\sum_{hk}\gamma_k\gamma_{hk}a_h+\sum_k\gamma_k^2-8\alpha\beta=0,$$

whence

$$\begin{aligned} \sum_k \gamma_{hk}^2 &= 8\alpha, & \sum_k \gamma_{hk}\gamma_{ik} &= 0,\ (h\neq i),\\ \sum_k \gamma_{hk}\gamma_k &= 0, & \sum_k \gamma_k^2-8\alpha\beta &= 0. \end{aligned} \tag{56}$$

When $\alpha=0$, all γ_{hk} and γ_k vanish, and (54) gives $x_1=x_2=x_3=0$, which is impossible. When $\alpha\neq 0$, the first line in (56) shows that by a

rotation of the flat space, we may make $\gamma_{hk} = 0$ for $h \neq k$, and suitable reflexions in the coördinate planes then give $\gamma_{hh} = -\sqrt{8\alpha}$, $(h = 1, 2, 3)$. The second line in (56) now gives $\gamma_1 = \gamma_2 = \gamma_3 = 0$, $\beta = 0$. Thus (55) is transformed into

$$R^2 c_k = \sqrt{8\alpha} \cdot a_k, \qquad (k = 1, 2, 3),$$
$$R^2 a = \alpha, \quad R^2 b = 2\sum_h a_h^2,$$

and substituting in (54), the terms independent of a_1, a_2, a_3 give

$$x_1^2 + x_2^2 + x_3^2 = \frac{1}{2}\alpha e^{4u_3}, \tag{57}$$

and the linear terms

$$\begin{aligned} 4x_1 &= \sqrt{8\alpha}\, e^{2u_3}(\mathrm{ch}^2 u_1 + \cos^2 u_2 - 2)(\mathrm{ch}^2 u_1 + \cos^2 u_2)^{-1}, \\ 4x_2 &= \sqrt{8\alpha}\, e^{2u_3} \cdot 2\,\mathrm{ch}\, u_1 \cos u_2 (\mathrm{ch}^2 u_1 + \cos^2 u_2)^{-1}, \\ 4x_3 &= \sqrt{8\alpha}\, e^{2u_3} \cdot 2\,\mathrm{sh}\, u_1 \sin u_2 (\mathrm{ch}^2 u_1 + \cos^2 u_2)^{-1}. \end{aligned} \tag{58}$$

By (58), $\partial x_i / \partial u_3 = 2x_i$, so that the coefficient of du_3^2 in $dx_1^2 + dx_2^2 + dx_3^2$ is $4(x_1^2 + x_2^2 + x_3^2) = 2\alpha e^{4u_3}$ by (57). On the other hand, (50) gives this coefficient as $e^{2\sigma_0} g = e^{4u_3}/4$ by (49) and (28), so that

$$8\alpha = 1. \tag{59}$$

Now it is seen that (58) is equivalent to (10), since (17) and (15) give

$$\begin{aligned} e^{2u_3}(\mathrm{ch}^2 u_1 + \cos^2 u_2)^{-1} &= \xi_3^2/(\xi_1^2 + \xi_2^2) = r_3^2/2, \\ \mathrm{ch}^2 u_1 + \cos^2 u_2 - 2 &= \xi_1^2 + \xi_2^2 - 2 = 2(r_1^2 + r_2^2 - r_3^2)/r_3^2, \\ 2\,\mathrm{ch}\, u_1 \cos u_2 &= 2\xi_1 \xi_2 = 2(r_1^2 - r_2^2)/r_3^2, \\ 2\,\mathrm{sh}\, u_1 \sin u_2 &= 2\sqrt{(\xi_1^2 - 1)(1 - \xi_2^2)} \\ &= (2/r_3^2)\sqrt{[(r_1 + r_2)^2 - r_3^2][r_3^2 - (r_1 - r_2)^2]}. \end{aligned}$$

DEPARTMENT OF PHYSICS, COLUMBIA UNIVERSITY.

ÜBER ADJUNGIERTE FUNKTIONALOPERATOREN.*

VON J. v. NEUMANN, PRINCETON.

Einleitung.

1. Diese Arbeit ist, wie mehrere frühere[1] den Operatoren des Hilbertschen Raumes gewidmet. Bezeichnungen und Begriffsbildungen lehnen an die in Anm. [1] genannten Arbeiten des Verfassers an, insbesondere wird die zu verwendende Terminologie in E. (S. 63—78) entwickelt. Demnach ist ein Operator A eine Funktion, die in einer Teilmenge des Hilbertschen Raumes $\mathfrak{H}$ definiert ist und Werte aus $\mathfrak{H}$ annimmt. Weitergehendes wird von Operatoren im allgemeinen nicht verlangt.

Wesentlich sind die folgenden Definitionen (vgl. a. a. O.):

I. Ein Operator A ist überall sinnvoll, wenn sein Definitionsbereich ganz $\mathfrak{H}$ ist.

II. A ist Fortsetzung von B, wenn der Definitionsbereich von B Teilmenge des Definitionsbereiches von A ist, und im ersteren A, B übereinstimmen, d. h. $Af = Bf$ gilt. Sind die Definitionsbereiche gleich, so ist $A = B$, sonst ist A echte Fortsetzung von B.

III. A ist linear, wenn mit $f_1, \cdots, f_n$ auch alle $a_1 f_1 + \cdots + a_n f_n$ zu seinem Definitionsbereiche gehören[2] und $A(a_1 f_1 + \cdots + a_n f_n) = a_1 Af_1 + \cdots + a_n Af_n$ gilt.

IV. A ist abgeschlossen, wenn für jede Folge $f_1, f_2, \cdots$ in seinem Definitionsbereiche, für die f_n gegen ein f und Af_n gegen ein f^* konvergiert $(n \to \infty)$, f auch zum Definitionsbereiche gehört[3] und $Af = f^*$ ist.

Die Operatoren, die wir betrachten werden, werden im allgemeinen nicht überall sinnvoll sein, jedoch wird ein gewisser minimaler Umfang des Definitionsbereiches allenfalls erwünscht sein, nämlich dieser:

---

* Received December 9, 1931.

[1] „Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren", Math. Annalen 102, 1 (1929), S. 49—131; „Zur Algebra der Funktionaloperationen und Theorie der normalen Funktionaloperatoren", Math. Annalen 102, 3 (1929), S. 370—427; „Zur Theorie der unbeschränkten Matrizen", Journal für Math., 161, 4 (1929), S. 208—236; „Über Funktionen von Funktionaloperatoren", Annals of Math. 32, 2 (1931), S. 191—226. Dieselben werden im Folgenden mehrfach zitiert werden, und zwar abgekürzt als E. bzw. A. bzw. M. bzw. F.

[2] D. h. der Definitionsbereich ist eine Linearmannigfaltigkeit.

[3] Daß der Definitionsbereich eine abgeschlossene Menge ist, würde hieraus nur folgen, wenn A stetig wäre. Stetigkeit folgte aus der Abgeschlossenheit nur, wenn $\mathfrak{H}$ kompakt wäre, was aber nicht der Fall ist. (Vgl. E., S. 70, insbesondere Anm. [29].)

V. A ist ein $*$-Operator, wenn die von seinem Definitionsbereiche aufgespannte, abgeschlossene Linearmannigfaltigkeit $\mathfrak{H}$ ist[4], d. h. wenn kein $f \neq 0$ zu seinem Definitionsbereiche orthogonal ist.

Wir betrachten nun die Bedingung

$$(f, Ag) = (f^*, g), \qquad *$$

wobei f, f^* als fest und g über den ganzen Definitionsbereich von A variabel gedacht sind. Bei gegebenem f mag es u. U. kein $*$ erfüllendes f^* geben, für gewisse f gibt es aber ein f^*: z. B. $f = 0$, $f^* = 0$. Daß nun $*$ im Lösbarkeitsfalle nur eine Lösung f^* besitze, ist für $*$-Operatoren charakteristisch[5], für diese und nur für diese kann daher die folgende Definition aufgestellt werden:

VI. Wenn A ein $*$-Operator ist, so ist A^* der folgende Operator: Sein Definitionsbereich besteht aus denjenigen f, für die $*$ lösbar ist, u. zw. ist dann $A^* f = f^*$.

Man erkennt sofort, daß A^*, welches in Anlehnung an die übliche Begriffsbildung, die es verallgemeinert[6], die Adjungierte von A heiße, linear und abgeschlossen ist. Ob es aber außer der 0 noch Stellen gibt, an denen es definiert ist, wissen wir im allgemeinen nicht, noch weniger, ob es ein $*$-Operator ist. Daß A^{**} (falls es gebildet werden kann, d. h. falls A^* ein $*$-Operator ist) Fortsetzung von A ist, ist evident.

Diese Begriffsbildungen erlauben, gewisse Begriffe aus E. einfacher zu umschreiben: so ist ein Hermitescher Operator A ein $*$-Operator, für den A^* Fortsetzung von A ist, ein hypermaximaler Operator einer mit $A^* = A$. (E., S. 70, Def. 6., sowie S. 72, Def. 8., 9.)[7].

2. Wir werden zeigen, daß die und nur die $*$-Operatoren A zu linear abgeschlossenen Operatoren B fortgesetzt werden können, für die auch A^* ein $*$-Operator ist. U. zw. gibt es dann unter allen diesen B einen kleinsten, d. h. einen, von dem alle anderen Fortsetzungen sind, er heiße $\tilde{A}$. Es ist $\tilde{A}^* = A^*$ und $\tilde{A} = A^{**}$. Ein weiteres Kriterium lautet so: Wenn eine Folge von Linearaggregaten $a_1^{(\nu)} f_1^{(\nu)} + \cdots + a_{\nu_n}^{(\nu)} f_{\nu_n}^{(\nu)}$ ($\nu = 1, 2, \cdots$, $f_1^{(\nu)}, \cdots, f_{\nu_n}^{(\nu)}$ aus dem Definitionsbereich) vorliegt derart, daß sie selbst

[4] Dies wurde in E., S. 70, Def. 6, für Hermitesche und in A., S. 405—406, Def. 5, 6, für normale Operatoren verlangt.

[5] Denn aus einer Lösung f^* gewinnt man alle anderen $f^* + h$, indem h über alle zum Definitionsbereich von A orthogonalen Elemente von $\mathfrak{H}$ läuft.

[6] Vgl. E., S. 111—112, insbesondere Anm. [64], sowie A., S. 372, Anm. [12], und S. 405, Def. 5.

[7] $A^* = A$ steht im Einklang mit der von M. Stone, Proc. Nat. Ac. 16 (1930), S. 173—174, für hypermaximale Operatoren verwendeten Bezeichnung „self-adjoint".

gegen 0 strebt, und $a_1^{(\nu)} Af_1^{(\nu)} + \cdots + a_{\nu_n}^{(\nu)} Af_{\nu_n}^{(\nu)}$ gegen f^* $(\nu \to \infty)$, so ist $f^* = 0$. Wenn A linear ist, lautet dies einfacher: Wenn $f_1, f_2, \cdots$ im Definitionsbereiche liegen und f_n gegen 0, Af_n gegen f^* strebt $(n \to \infty)$, so ist $f^* = 0$[8].

Diese Operatoren sind von besonderem Interesse:

VII. *A ist ein $**$-Operator, wenn A, A^* beide $*$-Operatoren sind.*

Wenn A ein $**$-Operator ist, so ist $\tilde{A}$, da es A fortsetzt und $\tilde{A}^* = A^*$ ist, ebenfalls $*$-Operator — d. h. $\tilde{A}$ ist linear abgeschlossen, $**$-Operator, Fortsetzung von A und hat dieselbe Adjungierte wie A. Wir gewinnen somit eine vollkommene Übersicht über alle $**$-Operatoren, wenn wir nur die linear abgeschlossenen unter ihnen betrachten. Diese sind wiederum, nach dem vorhin Gesagten, mit den linear abgeschlossenen $*$-Operatoren identisch.

Wir werden darum im folgenden stets linear abgeschlossene $*$-Operatoren betrachten, ist A ein solcher, so ist es nach dem vorhin Gesagten A^* ebenfalls, und es ist $A^{**} = A$ (da $\tilde{A} = A$ ist).

Betrachten wir nun den Operator A^*A, d. h. den folgendermaßen definierten: *wenn f im Definitionsbereiche von A und Af in dem von A^* liegt, ist A^*Af definiert, u. zw. gleich $A^*(Af)$.*

Auf Grund der Definition ist klar, daß $A^*A\,0$ Sinn hat und 0 ist, sowie die Linearität von A^*A; auch die Abgeschlossenheit ist leicht beweisbar[9]. Wenn A^*A ein $*$-Operator ist, ist es offenbar Hermitesch und definit[10]. Es ist aber vorderhand garnicht zu sehen, ob der Definitionsbereich von A^*A auch andere Elemente hat als 0, oder ob es gar ein $*$-Operator ist. Klar wäre dies nur, wenn A, A^* überall sinnvoll wären, was aber, wie wir an einen bekannten Satz von Toeplitz anlehnend zeigen werden, nur für stetiges A und A^* der Fall ist.

Es wird uns trotzdem gelingen, allgemein zu zeigen, daß A^*A ein $*$-Operator, ja sogar hypermaximal ist. Dasselbe gilt, wie Ersetzen von A durch A^* lehrt, für AA^*.

Es ist zweckmäßig, diese Verhältnisse durch einige Beispiele zu veranschaulichen. Realisieren wir den Hilbertschen Raum durch alle meß-

---

[8] Hieraus folgt die Stetigkeit nicht, weil die Existenz des Limes f^* vorausgesetzt wurde. Vgl. Anm. [3].

[9] Aus $f_n \to f$, $A^*Af_n \to f^*$ folgt: Für $m, n \to \infty$ ist $A^*Af_m - A^*Af_n \to 0$, also

$$(A^*Af_m - A^*Af_n,\ f_m - f_n) = (Af_m - Af_n,\ Af_m - Af_n) = \|Af_m - Af_n\|^2 \to 0,$$

also hat Af_m einen Limes, f'. Da A abgeschlossen ist, ist Af sinnvoll und gleich f', da A^* abgeschlossen ist, ist A^*f' sinnvoll und gleich f^* (man setze $Af_n = f_n'$). Also ist A^*Af sinnvoll und gleich f^*.

[10] $(A^*Af, f) = (Af, Af) = \|Af\|^2$ ist reell und $\geqq 0$, vgl. E., S. 70, Def. 6., und S. 74, Def. 10.

baren Funktionen $f(x)$ des Intervalles $0 \leqq x \leqq 1$ mit endlichem $\int_0^1 |f(x)|^2\,dx$ (vgl. z. B. E., S. 108—109); sei $\mathfrak{A}$ die Menge aller stetigen und differentiierbaren $f(x)$ mit endlichem $\int_0^1 |f'(x)|^2\,dx$, die gewisse, gleich anzugebende Randbedingungen erfüllen; A in $\mathfrak{A}$ definiert, u. zw. dort $Af = if'$. Als Randbedingungen lassen wir zu:

1. Keine Randbedingungen.
2. $f(1) : f(0) = C$ (C eine gegebene Zahl).
3. $f(1) = f(0) = 0$.

Entsprechend sprechen wir von $\mathfrak{A}_1, A_1$; $\mathfrak{A}_{2,C}, A_{2,C}$; $\mathfrak{A}_3, A_3$. $A_1, A_{2,C}, A_3$ sind linear, nicht abgeschlossen, $*$-Operatoren; A_1 wird von $A_{2,C}$ fortgesetzt, dieses von A_3, kein $A_{2,C}$ von einem anderen $A_{2,D}$. Aus

$$(Af, g) - (f, Ag) = i\int_0^1 (f'(x)\,\overline{g(x)} + f(x)\,\overline{g'(x)})\,dx$$
$$= i\,(f(x)\,\overline{g(x)})_0^1 = i\,(f(1)\,\overline{g(1)} - f(0)\,\overline{g(0)})$$

folgt, daß A_1^* Fortsetzung von A_3 ist, ebenso $A_{2,C}^*$ von $A_{2,1/\overline{C}}$, und A_3^* von A_1; also sind alle drei $*$-Operatoren, also $A_1, A_{2,C}, A_3$ $**$-Operatoren. Also setzen $A_1^* = \tilde{A}_1^*$, $A_{2,C}^* = \tilde{A}_{2,C}^*$, $A_3^* = \tilde{A}_3^*$ bzw. $\tilde{A}_3$, $\tilde{A}_{2,1/\overline{C}}$, $\tilde{A}_3$ fort. Man zeigt unschwer, worauf hier nicht näher eingegangen werde, daß diese bzw. Operatoren sogar gleich sind. Daraus folgt: $\tilde{A}_1$ ist Hermitesch, aber nicht hypermaximal; $\tilde{A}_{2,C}$ ist nur für $C = 1/\overline{C}$, d. h. $|C| = 1$, Hermitesch, aber dann auch hypermaximal; $\tilde{A}_3$ ist nicht Hermitesch. Dagegen sind die Operatoren $D_1 = \tilde{A}_1^*\,\tilde{A}_1 = \tilde{A}_3\,\tilde{A}_1$, $D_{2,C} = \tilde{A}_{2,C}^*\,\tilde{A}_{2,C} = \tilde{A}_{2,1/\overline{C}}\,\tilde{A}_{2,C}$, $D_3 = \tilde{A}_3^*\,\tilde{A}_3 = \tilde{A}_1\,\tilde{A}_3$ alle drei hypermaximal. Für alle drei ist $Af = -f''$, nur die Definitionsbereiche sind verschieden, die Randbedingungen lauten nämlich:

1. Für D_1: $f'(1) = f'(0) = 0$.
2. Für $D_{2,C}$: $f(1) : f(0) = C$, $f'(1) : f'(0) = 1/\overline{C}$.
3. Für D_3: $f(1) = f(0) = 0$.

Tatsächlich sind dies lauter zulässige Randbedingungen des Eigenwertproblems $-f'' = \lambda f$, was der Hypermaximalität gleichkommt (vgl. E., S. 49, Anm. [2], und S. 61).

Allgemeinere bzw. kompliziertere Beispiele (Sturm-Liouvillesche, partielle Differentialgleichungen usw.) zeigen dasselbe Verhalten.

3. Da A^*A und AA^* hermitesch, definit, hypermaximal sind, gibt es (wie sich zeigt) einen und nur einen hermiteschen, definiten, hypermaximalen

Operator B bzw. C mit $B^2 = A^*A$ bzw. $C^2 = AA^*$. Es zeigt sich, daß B denselben Definitionsbereich hat wie A, und stets $\| Bf\| = \| Af\|$ ist; dasselbe ist das Verhältnis von C und A^*. Sodann gelingt die Konstruktion eines Operators W, der im wesentlichen längentreu ist (vgl. w. u.), derart daß

$$A = WB = CW,\ A^* = BW^* = W^*C,\ C = WBW^*,\ B = W^*CW$$

ist. Wie man sieht, ist dies eine Zerlegung von A und A^*, wie sie der Zerlegung einer komplexen Zahl z und ihrer Konjugierten in einen reellen Faktor $\geqq 0$ (den Absolutwert) und einen Faktor vom Absolutwert 1 entspricht — die hypermaximalen Operatoren B, C benehmen sich ja in vielen Beziehungen wie reelle Zahlen, Definitität entspricht dem $\geqq 0$, Längentreue dem Absolutwert 1. Auffallend ist nur, daß zwei verschiedene Operatoren B, C den Absolutwert vertreten, allerdings gehen diese durch die längentreue Abbildung W ineinander über[11].

Für überall sinnvolle (also stetige, vgl. w. o.) Operatoren A, A^* ist diese Zerlegung in B, C, W trivial, es ist aber bemerkenswert, daß sie für alle linear abgeschlossenen $*$-Operatoren A erhalten bleibt und B, C immer hypermaximal ausfallen. Also wird z. B., wenn A hermitesch, definit, aber nicht maximal ist, also A^* echte Fortsetzung von A, also $A^*A = B^2$ Fortsetzung von A^2, B doch hypermaximal sein — also $B \neq A$, und B nicht etwa Fortsetzung von A, hat es doch denselben Definitionsbereich!

Auf Grund dieser Resultate ist es leicht zu zeigen, daß $B = C$ für die normalen Operatoren A, d. h. diejenigen, die eine komplexe Eigenwertdarstellung zulassen (vgl. A., S. 406, Def. 6., und S. 410—416, Satz 17—19), charakteristisch ist. D. h. $B^2 = C^2$, oder auch $A^*A = AA^*$. Diese Bedingung ist bekanntlich im endlichvieldimensionalen Raume[12] sowie bei vollstetigen Operatoren des Hilbertschen Raumes[13] für die Normalität charakteristisch, im Hilbertschen Raume zeigte der Verf. noch, daß sie für alle stetigen Operatoren die charakteristische ist (E., S. 115—116, sowie A., S. 406, und der w. o. a. O.), für allgemeine A gewann er aber a. a. O. eine anders gebaute Bedingung. Jetzt ist $A^*A = AA^*$ als allgemein charakteristisch erkannt, ja, es genügt, wenn A^*A Fortsetzung von AA^* ist oder umgekehrt, da beide Operatoren hypermaximal sind.

4. Wenn A irgendein linearer Operator ist, so können wir in seinem Definitionsbereiche $\mathfrak{A}$ durch die Definition $\| f\|_A = \| Af\|$ eine neue Metrik

[11] Für vollstetige Operatoren untersuchte E. Schmidt die zu B, C gehörigen Eigenwertprobleme, vgl. Math. Ann. 63 (1907), S. 433—476, §§ 14—16.

[12] Frobenius, Journ. f. Math., 84 (1877), S. 51—54.

[13] Hellinger und Toeplitz, Enzykl. d. Math. Wiss., II. C. 13., S. 1562—1563.

einführen. Damit $\mathfrak{A}$ in $\mathfrak{H}$ überall dicht sei, muß A ein $*$-Operator sein, damit es im Sinne der neuen Metrik topologisch vollständig sei (wenigstens relativ zu $\mathfrak{H}$[14]), muß A abgeschlossen sein[15] — d. h. wir werden zu unserer Operatorenklasse (linear, abgeschlossen, *) zurückgeführt.

A_1, A_2 bestimmen dieselbe Metrik, wenn sie denselben Definitionsbereich haben, und dort $\| A_1 f \| = \| A_2 f \|$ gilt — wir nennen sie dann metrisch gleich, $A_1 \underset{m}{=} A_2$. Ist der Definitionsbereich von A_1 Teil desjenigen von A_2 und im ersteren $\| A_1 f \| = \| A_2 f \|$, so nennen wir A_2 eine metrische Fortsetzung von A_1. Sind die Definitionsbereiche gleich, so ist $A_1 \underset{m}{=} A_2$, sonst sei A_2 echte metrische Fortsetzung von A_1.

Wie wir sahen, ist jedes A einem hermiteschen, definiten, hypermaximalen B metrisch gleich, und man zeigt leicht, daß B hierdurch eindeutig bestimmt ist (vgl. w. u.).

Andererseits werden wir zeigen, daß jeder unstetige Operator A (auch wenn er im gewöhnlichen Sinne unfortsetzbar, d. h. maximal ist), echte metrische Fortsetzungen besitzt. Hier gibt es also keine Maximalität!

5. Unsere Methode wird im wesentlichen darauf beruhen, den Raum aller Paare $\{f, g\}$ (f, g beliebig aus $\mathfrak{H}$) zu betrachten. Wenn wir

$$a\{f, g\} = \{af, ag\},$$
$$\{f, g\} + \{h, k\} = \{f+h, g+k\},$$
$$(\{f, g\}, \{h, k\}) = (f, h) + (g, k)$$

definieren (also $\| \{f, g\} \|^2 = \| f \|^2 + \| g \|^2$), so ist dies ein Hilbertscher Raum, denn die charakteristischen Bedingungen A.—E. aus E., S. 64—66, gelten für ihn genau so wie für $\mathfrak{H}$. Er verhält sich zu $\mathfrak{H}$ etwa so, wie die Ebene zur Geraden, wir wollen ihn H nennen. Jeder Operator A (in $\mathfrak{H}$) kann in H durch die Menge aller Paare $\{f, Af\}$ dargestellt werden — dies ist das Analogon der graphischen Darstellung der auf der Zahlengeraden definierten Funktionen in der Ebene. Die genannte Menge heiße das Bild von A, B_A.

Man erkennt sofort: $A = B$ bedeutet $B_A = B_B$, A Fortsetzung von B bedeutet B_A umfaßt (als Menge) B_B, A linear bedeutet, daß B_A eine Linearmannigfaltigkeit ist, A abgeschlossen bedeutet, daß B_A (als Menge) abgeschlossen ist.

---

[14] D. h. daß $\| f_m - f_n \|_A \to 0$ $(m, n \to \infty)$ die Existenz eines $\| \cdots \|_A$-Limes von $f_1, f_2, \cdots$ in $\mathfrak{A}$ nach sich zieht — soweit ein $\| \cdots \|$ Limes in $\mathfrak{H}$ existiert. Wenn stets $\| Af \| \geqq \varepsilon \cdot \| f \|$ für ein festes $\varepsilon > 0$ gilt, ist die letztere Beschränkung überflüssig.

[15] Diese Betrachtung verschiedener Metrisierungen in $\mathfrak{H}$, so wie der topologischen Vollständigkeit wurde von Herrn K. Friedrichs in einem Vortrage angeregt, den er im Juni 1931 in der Göttinger Math. Ges. hielt.

Wir werden unsere Resultate durch Untersuchungen der B_A in H gewinnen. Die ganze Operatorentheorie ließe sich sogar noch wesentlich konsequenter in H betreiben, als wie es hier geschieht, doch gehen wir hierauf nicht ein.

I. Die Operatoren A, A^*, A^*A.

1. Wir untersuchen, wie in Einl. 4 auseinandergesetzt wurde, den Hilbertschen Raum H und die Bilder B_A in H an Stelle der Operatoren A in $\mathfrak{H}$. Wir führen noch den Operator

$$\boldsymbol{U}\{f, g\} = \{ig, -if\}$$

in H ein, er ist Hermitesch und unitär, d. h. $\boldsymbol{U} = \boldsymbol{U}^* = \boldsymbol{U}^{-1}$ (also $\boldsymbol{U}^2 = 1$). Da $(f, Ag) = (f^*, g)$ als $(\{f, f^*\}, \boldsymbol{U}\{g, Ag\}) = 0$ geschrieben werden kann, bedeutet $A^* f = f^*$, daß $\{f, f^*\}$ zu $\boldsymbol{U}B_A$ orthogonal ist, d. h. B_{A^*} ist die Menge aller zu $\boldsymbol{U}B_A$ orthogonalen Elemente von H — d. h. eine abgeschlossene Linearmannigfaltigkeit, also A^* linear abgeschlossen.

Daß es linear abgeschlossene Fortsetzungen von A gibt, bedeutet, daß es abgeschlossene Linearmannigfaltigkeiten B_B gibt, die B_A enthalten. Nun kann jede Menge in H ein B_B sein, wenn sie nie zwei $\{f, f_1^*\}$, $\{f, f_2^*\}$ mit $f_1^* \neq f_2^*$ enthält — oder, was dasselbe ist, weil es sich um Linearmannigfaltigkeiten handelt, wenn sie nie $\{0, f^*\}$ mit $f^* \neq 0$ enthält. Es ist also bloß zu fordern: die kleinste B_A enthaltende, d. h. die von B_A aufgespannte abgeschlossene Linearmannigfaltigkeit soll kein $\{0, f^*\}$ mit $f^* \neq 0$ enthalten. Ersetzen wir diese Begriffe durch ihre Definitionen, so wird:

SATZ 1. *Der *-Operator A mit dem Definitionsbereich $\mathfrak{A}$ hat dann und nur dann eine linear abgeschlossene Fortsetzung B, wenn Folgendes gilt: Liegen alle $f_1^{(n)}, \cdots, f_{\nu_n}^{(n)}$ $(n = 1, 2, \cdots)$ in $\mathfrak{A}$, und gilt*

$$a_1^{(n)} f_1^{(n)} + \cdots + a_{\nu_n}^{(n)} f_{\nu_n}^{(n)} \to 0, \qquad a_1^{(n)} A f_1^{(n)} + \cdots + a_{\nu_n}^{(n)} f_{\nu_n}^{(n)} \to f^* \quad (n \to \infty),$$

so ist $f^* = 0$[16]*. U. zw. gibt es dann unter diesen B ein kleinstes, von dem alle anderen Fortsetzungen sind, nämlich dasjenige, für welches B_B die von B_A aufgespannte Linearmannigfaltigkeit ist.*

Ein weiteres Kriterium entsteht so. Wenn A^* auch ein *-Operator ist, so existiert auch A^{**}, es ist linear abgeschlossen und offenbar Fortsetzung von A, also existiert $\tilde{A}$. Existiert umgekehrt $\tilde{A}$, so ist $B_{\tilde{A}}$ die von B_A aufgespannte abgeschlossene Linearmannigfaltigkeit, entsprechend $\boldsymbol{U}B_{\tilde{A}}$ und $\boldsymbol{U}B_A$,

[16] Die hier stets verwendete starke Konvergenz könnte auch durch schwache ersetzt werden (vgl. z. B. A., S. 378–381), denn für Linearmannigfaltigkeiten (d. i. B_A, B_B) ist die starke Abgeschlossenheit der schwachen, wie E. Schmidt zeigte (Rend. d. Circ. Mat. d. Palermo, 25 (1908), S. 57–73), gleichwertig.

also sind zum ersteren dieselben Elemente von H orthogonal wie zum letzteren — d. h. $\tilde{A}^* = A^*$. Ferner ist $B_{\tilde{A}^*} = B_{A^*}$ die Komplementäre von $\boldsymbol{U}B_{\tilde{A}}$ (es handelt sich jetzt um zwei abgeschlossene Linearmannigfaltigkeiten, vgl. E. S. 74), also, da $\boldsymbol{U}$ unitär, $\boldsymbol{U}^2 = 1$, und die Komplementarität eine symmetrische Beziehung ist, $B_{\tilde{A}}$ die Komplementäre von $\boldsymbol{U}B_{A^*}$. Das bedeutet in der Terminologie von Einl. 1, daß die dortige Def. VI für A^{**} einen eindeutigen Operator, nämlich $\tilde{A}$ ergibt. Also ist auch A^* ein *-Operator, und $A^{**} = \tilde{A}$. Wir haben bewiesen:

Satz 2. *Für die Existenz von $\tilde{A}$ ist auch dieses charakteristisch: A^* ist auch ein $*$-Operator (für A war das Voraussetzung). U. zw. ist dann $\tilde{A}^* = A^*$, $\tilde{A} = A^{**}$.*

Man beachte, daß zwischen den Definitionsbereichen von A und A^* keinerlei Zusammenhänge gefordert oder behauptet wurden.

2. Sei nunmehr A ein linear abgeschlossener $*$-Operator, dann ist es A^* nach Satz 2 ebenfalls. Da $\boldsymbol{U}B_A$, B_{A^*} in H komplementär sind, kann jedes Element in H auf eine und nur eine Weise als Summe eines Elementes von $\boldsymbol{U}B_A$ und eines von B_{A^*} dargestellt werden. (vgl. z. B. E., S. 74, Satz 13). Wir wählen das H-Element als $\{0, -if\}$, dann ist

$$\{0, -if\} = \{iAg, -ig\} + \{h, A^*h\}$$

eindeutig lösbar (g im Definitionsbereiche von A, h in dem von A^*). D. h. $iAg + h = 0$, $-ig + A^*h = -if$. Also ist Ag gleich $-ih$, und somit ebenfalls im Definitionsbereiche von A^*, und dann

$$-if = -ig + A^*h = -i(g + A^*Ag) = -i(A^*A + 1)g,$$
$$(A^*A + 1)\, g = f.$$

Der Operator $A^*A + 1$ ist also für so viele g sinnvoll, daß ihr Wertevorrat der ganze Hilbertsche Raum ist!

Der inverse Operator D, $Df = g$, ist also linear und überall definiert. Ferner ist

$$(f, Df) = (A^*Ag + g, g) = (A^*Ag, g) + (g, g)$$
$$= (Ag, Ag) + (g, g) = \|Ag\|^2 + \|g\|^2$$

erstens reell, also D Hermitesch. Zweitens ist es $\geqq 0$, und nur $= 0$ für $\|g\|^2 = 0$, $g = 0$, $f = (A^*A + 1)g = 0$, also D definit und $Df \neq 0$ für $f \neq 0$. Drittens ist es $\geqq \|g\|^2 = \|Df\|^2$, also

$$\|Df\|^2 \leqq (f, Df) \leqq \|f\| \cdot \|Df\|, \quad \|Df\| \leqq \|f\|,$$

also D stetig. Zusammenfassend: D ist ein Hermitescher, definiter, stetiger Operator, aus $f \neq 0$ folgt $Df \neq 0$, das ganze Spektrum von D liegt

zwischen 0 und 1[17], und es ist $A^*A+1 = D^{-1}$. Hieraus ist es ein Leichtes, die Hypermaximalität von A^*A+1 zu folgern, z. B. so. Daß f zu allen Dk orthogonal ist, bedeutet, daß Df zu allen k orthogonal ist (D ist Hermitesch und überall sinnvoll), also $Df = 0$, $f = 0$ — die Komplementäre der vom Wertevorrat von D aufgespannten abgeschlossenen Linearmannigfaltigkeit (in $\mathfrak{H}$) ist (0), also die letztere $\mathfrak{H}$. Da der Wertevorrat von D der Definitionsbereich von A^*A+1 ist, also auch der von A^*A, ist A^*A ein $*$-Operator. Nach Einl. 2, insbesondere Anm. [10], ist also A^*A Hermitesch und definit, daher ist es A^*A+1 auch. Da A^*A+1 alle Werte in $\mathfrak{H}$ annimmt, ist E., S. 102, Satz 41 anwendbar: A^*A+1 ist hypermaximal, also auch A^*A. Also:

SATZ 3. *A sei ein linear abgeschlossener $*$-Operator. Dann ist es A^* auch, und A^*A ist Hermitesch, definit und hypermaximal.*

3. Der Definitionsbereich $\mathfrak{A}_1$ von A^*A ist Teil desjenigen $\mathfrak{A}$ von A, das auf $\mathfrak{A}_1$ eingeschränkte A heiße A_1 — A_1 ist offenbar linear und $*$, aber nicht notwendig abgeschlossen. Da A Fortsetzung von A_1 ist, ist A_1^* Fortsetzung von A^*, also auch $*$, so daß A_1 sogar $**$ ist. Wir wollen $\tilde{A}_1 = A$ beweisen, d. h. daß B_A die von B_{A_1} aufgespannte abgeschlossene Linearmannigfaltigkeit (in H) ist. Anders formuliert: Die Differenz von B_A und der von B_{A_1} aufgespannten abgeschlossenen Linearmannigfaltigkeit ist (0), oder: in B_A ist nur 0 zur genannten Menge, oder, was dasselbe bedeutet, zu B_{A_1}, orthogonal (vgl. E., S. 74, Def. 11). Ein Element von B_A ist ein (k, Ak), wenn es zu B_{A_1} orthogonal ist, so ist es insbesondere zu den am Anfang von I., 2. konstruierten $\{g, Ag\}$ orthogonal, also $\{iAk, -ik\}$ zu $\{iAg, -ig\}$ (denn A^*Ag war sinnvoll, als gehört g zu $\mathfrak{A}_1$). Zum dortigen $\{h, A^*h\}$, das zu B_{A^*} gehört, ist das zu UB_A gehörige $\{iAk, -ik\}$ erst recht orthogonal, also auch zu deren Summe, $\{0, -if\}$. Daß $\{iAk, -ik\}$ und $\{0, -if\}$ orthogonal sind (in H), bedeutet, daß es k, f sind (in $\mathfrak{H}$), u. zw. für jedes f — also ist $k = 0$, $\{k, Ak\} = \{0, 0\}$. Damit sind wir am Ziele:

SATZ 4. *Der Definitionsbereich von A sei $\mathfrak{A}$, der von A^*A sei $\mathfrak{A}_1$ (Teil von $\mathfrak{A}$), das durch $\mathfrak{A}_1$ eingeschränkte A heiße A_1. A_1 ist linear, $**$, aber nicht notwendig abgeschlossen. Dagegen ist $\tilde{A}_1 = A$.*

4. Wenn B ein Hermitescher hypermaximaler Operator ist, so gehört nach E., S. 91, Def. 17 und S. 92, Satz 36 eine Zerlegung der Einheit $E(\lambda)$ dazu. Wenn $E(\lambda)$ für $\lambda < 0$ konstant (also $= 0$) ist, so ist nach den dortigen Formeln stets $(Bf, f) \geqq 0$, d. h. B definit; ist dagegen $E(\lambda)$ für ein $\lambda < 0$ nicht 0, so gibt es für diese $\lambda = \lambda_0 < 0$ ein $f \neq 0$ mit $E(\lambda_0)f = f$,

[17] Letzteres wegen $0 \leqq (f, Df) \leqq \|f\|^2$, woraus es durch ähnliche Rechnungen wie A., S. 418, folgt.

also $E(\lambda)f = f$ für $\lambda \geqq \lambda_0$ (wegen $E(\lambda_0)$ Teil von $E(\lambda)$ vgl., das in E. über Projektionsoperatoren Gesagte), und dann errechnet man auf Grund derselben Formeln $(Bf, f) \leqq \lambda_0 \| f \|^2 < 0$, d. h. B ist nicht definit. Für die Definititität von B ist also $E(\lambda)$ konstant (also $= 0$) für $\lambda < 0$ charakteristisch.

Betrachten wir nun die Schar

$$F(\mu)\begin{cases} = 0 \text{ für } \mu < 0, \\ = E(\sqrt{\lambda}) \text{ für } \mu \geqq 0. \end{cases}$$

Auch sie ist offenbar eine Zerlegung der Einheit, und der dazugehörige Hermitesche hypermaximale Operator B' ist definit. Wenn $B'f$ Sinn hat, so ist nach Definition

$$\int \mu^2 d \| F(\mu) f \|^2 = \int \mu^2 d \| E(\sqrt{\mu}) f \|^2 = \int \lambda^4 d \| E(\lambda) f \|^2$$

endlich, also auch $\int \lambda^2 d \| E(\lambda) f \|^2$ $\left(\text{denn } \int d \| E(\lambda) f \|^2 \text{ ist es ebenfalls und } \lambda^2 \leqq \tfrac{1}{2} + \tfrac{1}{2}\lambda^4\right)$—also hat Bf Sinn. Dieselbe Rechnung, wie sie in F., S. 206 ausgeführt wurde (um die dortige Formel e) zu beweisen), zeigt, daß

$$\int \lambda^2 d \| E(\lambda) Bf \|^2 = \int \lambda^4 d \| E(\lambda) f \|^2$$

ist, also endlich—also hat auch $B(Bf)$, d. h. $B^2 f$ Sinn. Und wieder eine entsprechende Rechnung zeigt, daß

$$(B^2 f, g) = (B(Bf), g) = \int \lambda \, d(E(\lambda) Bf, g) = \int \lambda^2 d(E(\lambda) f, g)$$
$$= \int \mu \, d(E(\sqrt{\mu}) f, g) = \int \mu \, d(F(\mu) f, g) = (B'f, g)$$

ist, d. h. $B^2 f = B'f$. Daher ist B^2 Fortsetzung des hypermaximalen B', also $B^2 = B'$.

Etwas anders angeordnet, können wir dies auch so formulieren:

Satz 5. *B, B' seien Hermitesche, hypermaximale, definite Operatoren, die bzw. zu ihnen gehörigen Zerlegungen der Einheit $E(\lambda)$, $F(\lambda)$—also $E(\lambda) = F(\lambda) = 0$ für $\lambda < 0$. Dann ist $B^2 = B'$ mit $E(\lambda) = F(\lambda^2)$ für alle $\lambda \geqq 0$ gleichbedeutend.*

Wenn also B' gegeben ist und B nicht, so hat $B^2 = B'$ eine einzige derartige Lösung B, sie heiße $\sqrt{B'}$.

Wir wenden nun Satz 5 auf $B' = A^*A$ an und setzen $B = \sqrt{A^*A}$. Der Definitionsbereich von $B^2 = A^*A$ hieß $\mathfrak{A}_1$, das hierauf eingeschränkte A

A_1, das hierauf eingeschränkte B heiße B_1 (der Definitionsbereich von B umfaßt den von B^2, d. h. $\mathfrak{A}_1$). Nach Satz 3 ist $\tilde{A}_1 = A$, und wenn wir darin A durch B ersetzen (wegen der Hypermaximalität ist $B^* = B$), $\tilde{B}_1 = B$.

A_1, B_1 haben denselben Definitionsbereich, und in ihm ist

$$\|A_1 f\|^2 = (A_1 f,\ A_1 f) = (Af,\ Af) = (A^* Af,\ f)$$
$$\|B_1 f\|^2 = (B_1 f,\ B_1 f) = (Bf,\ Bf) = (B^2 f,\ f),$$

d. h. $\|A_1 f\| = \|B_1 f\|$. Also sind A_1, B_1 metrisch gleich (vgl. Einl. 4.), und daraus folgt die metrische Gleichheit von $\tilde{A}_1$, $\tilde{B}_1$[18], d. h. von A, B. Einsetzen in die Definition der metrischen Gleichheit ergibt:

SATZ 6. *Sei A wie in Satz 2., 3., $B = \sqrt{A^* A}$ (vgl. Satz 5.), also Hermitesch, definit, hypermaximal. Dann sind A, B metrisch gleich, d. h.: beide haben denselben Definitionsbereich $\mathfrak{A}$, und in ihm ist stets $\|Af\| = \|Bf\|$.*

5. Nennen wir den Wertevorrat von A $\mathfrak{W}$, und die abgeschlossenen Hüllen der Linearmannigfaltigkeiten $\mathfrak{A}$, $\mathfrak{W}$ bzw. $\overline{\mathfrak{A}} = \mathfrak{H}$, $\overline{\mathfrak{W}}$. Ferner heiße die Menge aller f mit $Af = 0$ $\mathfrak{N}$, sie ist Teil von $\mathfrak{A}$ (und $\overline{\mathfrak{A}}$), und da A linear abgeschlossen ist, eine abgeschlossene Linearmannigfaltigkeit (alles in $\mathfrak{H}$). Die entsprechenden Mengen für A^* seien $\mathfrak{A}'$, $\mathfrak{W}'$, $\overline{\mathfrak{A}'} = \mathfrak{H}$, $\overline{\mathfrak{W}'}$, $\mathfrak{N}'$.

f orthogonal zu $\overline{\mathfrak{W}}$ bedeutet dasselbe wie f orthogonal zu $\mathfrak{W}$, d. h. zu allen Ag, d. h. stets $(f, Ag) = 0$ — und dies besagt, daß $A^* f$ definiert ist und $= 0$, d. h. f in $\mathfrak{N}'$. Also: $\mathfrak{N}' = \mathfrak{H} - \overline{\mathfrak{W}}$. Ebenso ist: $\mathfrak{N} = \mathfrak{H} - \overline{\mathfrak{W}'}$. Zusammenfassend:

$$\mathfrak{N} \subset \mathfrak{A},\ \mathfrak{N} = \mathfrak{H} - \overline{\mathfrak{W}'},\ \mathfrak{N}' \subset \mathfrak{A}',\ \mathfrak{N}' = \mathfrak{H} - \overline{\mathfrak{W}}.$$

Da die Menge der $Af = 0$ nach Satz 4. mit derjenigen der $Bf = 0$ identisch ist, ist $\mathfrak{N}$ dieselbe auch für B. Ist der Wertevorrat von B $\mathfrak{V}$ (Linearmannigfaltigkeit!) und seine abgeschlossene Hülle $\overline{\mathfrak{V}}$ (abgeschlossene Linearmannigfaltigkeit!), so ergibt dieselbe Betrachtung $\mathfrak{N} = \mathfrak{H} - \overline{\mathfrak{V}}$ ($B^* = B$ wegen der Hypermaximalität), d. h. $\overline{\mathfrak{V}} = \overline{\mathfrak{W}'}$.

Ersetzen wir A durch A^*, also A^* durch $A^{**} = A$, so entstehen nach Satz 5., 6. die Hermiteschen, definiten, hypermaximalen Operatoren AA^* und $C = \sqrt{AA^*}$. Nach Satz 6. sind A^*, C metrisch gleich, d. h. beide haben denselben Definitionsbereich $\mathfrak{A}'$, und in ihm ist stets $\|A^* f\| = \|Cf\|$.

---

[18] C sei ein linearer $**$-Operator (A_1, B_1 sind es nach Satz 4.), dann ist B_C eine Linearmannigfaltigkeit, also $B_{\tilde{C}}$ seine abgeschlossene Hülle. D. h. $\tilde{C}$ ist so definiert: Wenn $f_1, f_2, \cdots$ im Definitionsbereiche von C liegen, $f_n \to f$ $(n \to \infty)$ und auch die Cf_n einen Limes haben — d. h. $\|C(f_m - f_n)\| \to 0$ $(m, n \to \infty)$ — so liegt f im Definitionsbereiche von $\tilde{C}$, und $\tilde{C}f$ ist der Limes der Cf_n. Diese Formulierung macht es klar, daß beim Ersetzen von C durch einen metrisch gleichen Operator auch $\tilde{C}$ in einen metrisch gleichen übergeht.

Wir bilden auch für C die Mengen $\mathfrak{V}'$ (Wertevorrat), $\overline{\mathfrak{V}}'$, es ist $\overline{\mathfrak{V}}' = \overline{\mathfrak{W}}$. Die Zuordnung $Bf \Longleftrightarrow Af$ bildet nach Satz 4. $\mathfrak{V}$ eindeutig, linear und längentreu auf $\mathfrak{W}$ ab, sie kann daher eineindeutig, linear und längentreu auf die abgeschlossenen Hüllen $\overline{\mathfrak{V}} = \overline{\mathfrak{W}}'$, $\overline{\mathfrak{W}}$, d. h. $\mathfrak{H} - \mathfrak{N}$, $\mathfrak{H} - \mathfrak{N}'$ ausgedehnt werden. Der Operator dieser längentreuen Abbildung heiße V, sein Definitionsbereich ist $\mathfrak{H} - \mathfrak{N}$, sein Wertevorrat $\mathfrak{H} - \mathfrak{N}'$ (vgl. E., S. 70, Def. 6. und S. 71, Satz 10.). Eine ebensolche Abbildung von $\mathfrak{V}'$ auf $\mathfrak{W}'$ ist $Cf \Longleftrightarrow A^*f$, sie kann ebenso auf $\overline{\mathfrak{V}}' = \overline{\mathfrak{W}}$, $\overline{\mathfrak{W}}'$, d. h. $\mathfrak{H} - \mathfrak{N}'$, $\mathfrak{H} - \mathfrak{N}$ ausgedehnt werden. Ihr längentreuer Operator heiße V', sein Definitionsbereich ist $\mathfrak{H} - \mathfrak{N}'$, sein Wertevorrat $\mathfrak{H} - \mathfrak{N}$. Auf Grund der Definition von V, V' ist

$$A = VB, \quad A^* = V'C.$$

Da es gleichgültig ist, wie V bzw. V' außerhalb des Wertevorrates von B bzw. C, d. h. von $\mathfrak{H} - \mathfrak{N}$ bzw. $\mathfrak{H} - \mathfrak{N}'$ definiert werden, können wir sie zu überall definierten linearen Operatoren W bzw. W' fortsetzen, die in $\mathfrak{N}$ bzw. $\mathfrak{N}'$ verschwinden. Dann ist wieder

$$A = WB, \quad A^* = W'C.$$

Wenn wir die Projektionsoperatoren von $\mathfrak{H} - \mathfrak{N}$ bzw. $\mathfrak{H} - \mathfrak{N}'$ E bzw. E' nennen, so ist

$$W = VE, \quad W' = V'E',$$

woraus sofort

$$\|Wf\| = \|Ef\|, \quad \|W'f\| = \|E'f\|$$

folgt, also die Stetigkeit von W, W'. Ferner folgt hieraus

$$\text{◎} \qquad W^*W = E, \quad W'^*W' = E'.^{19}$$

Da W' eine Abbildung $\mathfrak{H} - \mathfrak{N} \Longleftrightarrow \mathfrak{H} - \mathfrak{N}'$ vermittelt, und $W^*W = E$ die identische Abbildung $\mathfrak{H} - \mathfrak{N} \Longleftrightarrow \mathfrak{H} - \mathfrak{N}$, vermittelt W^* die inverse, und ebenfalls längentreue Abbildung $\mathfrak{H} - \mathfrak{N}' \Longleftrightarrow \mathfrak{H} - \mathfrak{N}$. Daher ist in $\mathfrak{H} - \mathfrak{N}'$ $\|W^*f\| = \|f\|$. In $\mathfrak{N}'$ ist, da es zum Wertevorrat von W, $\mathfrak{H} - \mathfrak{N}'$, orthogonal ist, $W^*f = 0$ — daher gilt allgemein $W^*E'f = W^*f$, also $\|W^*f\| = \|E'f\|$. Hieraus folgt (vgl. Anm. [19])

$$\text{◎} \qquad WW^* = E',$$

[19] Es ist

$$\|Wf\|^2 = (Wf, Wf) = (W^*Wf, f), \qquad \|Ef\|^2 = (Ef, f),$$

also $W^*W = E$; ebenso $W'^*W' = E'$.

ebenso ist

$$\text{◎} \qquad W' W'^* = E.$$

Linksmultiplikation von $A = WB$ mit W^* ergibt, da $EB = B$ ist (B hat Werte aus $\mathfrak{H} - \mathfrak{R}$),

$$* \qquad B = W^* A,$$

ebenso ergibt Linksmultiplikation von $A^* = W' C$ mit W'^*

$$\divideontimes \qquad C = W'^* A^*.$$

Wenden wir auf die vier $\divideontimes$-Formeln $*$ an, so zeigt eine einfache Diskussion

$$\# \qquad A^* = BW^*, \qquad A = CW'^*,$$
$$\# \qquad B = A^* W, \qquad C = AW'.$$

Bilden wir nun den Operator WBW^*[20]. Da W überall sinnvoll ist, ist er sinnvoll, wenn $W^* f$ in $\mathfrak{A}$ liegt. Da $\mathfrak{R}$ Teil von $\mathfrak{A}$ ist, gehört für jedes Element von $\mathfrak{A}$ seine $\mathfrak{R}$-Projektion zu $\mathfrak{A}$, also auch seine $\mathfrak{H} - \mathfrak{R}$-Projektion — d. h. der in $\mathfrak{H} - \mathfrak{R}$ liegende Teil von $\mathfrak{A}$ ist dort überall dicht, d. h. der im Wertevorrat von W^* (dieser ist $\mathfrak{H} - \mathfrak{R}$, vgl. w. o.) liegende Teil von $\mathfrak{A}$ ist dort überall dicht. Da W^* eine Abbildung $\mathfrak{H} - \mathfrak{R}' \Longleftrightarrow \mathfrak{H} - \mathfrak{R}$ vermittelt, liegen in $\mathfrak{H} - \mathfrak{R}'$ die f mit $W^* f$ aus $\mathfrak{A}$ überall dicht. Für f in $\mathfrak{R}'$ ist $W^* f = 0$, also auch in $\mathfrak{A}$ — also liegen die genannten f überall dicht, WBW^* ist ein $*$-Operator. Hermitesch und definit ist er ebenso wie B. Auch hypermaximal ist WBW^*, denn sei für alle g, für die es Sinn hat

$$(f, WBW^* g) = (f^*, g),$$

d. h.

$$(W^* f, BW^* g) = (f^*, g).$$

Sei $g = Wh$, h aus $\mathfrak{H} - \mathfrak{R}$, Bh sinnvoll, dann ist $Wg = W^* Wh = h$,

$$(W^* f, Bh) = (f^*, Wh) = (W^* f^*, h).$$

Für jedes h aus $\mathfrak{R}$ hat Bh Sinn, u. zw. 0, und auch rechts steht, da $W^* f^*$ zu $\mathfrak{H} - \mathfrak{R}$ gehört, 0. Also ist stets

$$(W^* f, Bh) = (W^* f^*, h).$$

[20] Operatorenprodukte fassen wir konsequent so auf: RSf ist definiert, wenn Sf und $R(Sf)$ definiert sind, und ist gleich $R(Sf)$; entsprechend für QRS, $PQRS$, ···. Diese Multiplikation ist offenbar associativ.

Da B hypermaximal ist, ist BW^*f sinnvoll und gleich W^*f^*, also $WBW^*f = WW^*f^* = f^*$, falls f^* zu $\mathfrak{N}'$ gehört. Dieses folgt aber aus der ursprünglichen Gleichung, da für alle g von $\mathfrak{N}'$ $W^*g = 0$ ist, also deren linke Seite 0 ist.

Nun ist (man beachte die $\ast$-, ◎-, #-Formeln!)[20]

$$\begin{aligned}(WBW^*)^2 &= WBW^* \cdot WBW^* = WB \cdot W^*W \cdot BW^* \\ &= WB \cdot E \cdot BW^* = WB \cdot EB \cdot W^* \\ &= WB \cdot BW^* = A \cdot A^* = C^2,\end{aligned}$$

also nach Satz 4 $WBW^* = C$. Nach # besagt dies $WA^* = C$, und da ein $\ast$ $W'^*A^* = C$ aussagt, $WA^* = W'^*A^*$. Somit gilt $Wf = W'^*f$ im ganzen Wertevorrat von A^*, d. i. in $\mathfrak{W}'$ — also auch in $\overline{\mathfrak{W}}' = \mathfrak{H} - \mathfrak{N}$. In $\mathfrak{N}$ gilt es aber auch, da dort beide Seiten gleich 0 sind, also ist $W = W'^*$; Anwenden von $*$ ergibt $W^* = W'$.

Zusammenfassend haben wir also:

SATZ 7. *Sei A wie in Satz 2, 3, $B = \sqrt{A^*A}$, $C = \sqrt{AA^*}$ (vgl. Satz 4 mit A und mit A^*), also hermitesch, definit, hypermaximal. A, A^*, B, C mögen die bzw. Linearmannigfaltigkeiten $\mathfrak{A}$, $\mathfrak{A}'$, $\mathfrak{A}$, $\mathfrak{A}'$ als Definitionsbereiche, $\mathfrak{W}$, $\mathfrak{W}'$, $\mathfrak{B}$, $\mathfrak{B}'$ als Wertevorräte haben, – deute ihre bzw. abgeschlossenen Hüllen an. Die bzw. Mengen (abgeschlossenen Linearmannigfaltigkeiten) $Af = 0$, $A^*f = 0$, $Bf = 0$, $Cf = 0$ seien $\mathfrak{N}$, $\mathfrak{N}'$, $\mathfrak{N}$, $\mathfrak{N}'$. Es ist:*

$$\mathfrak{N} \subseteq \mathfrak{A} \subseteq \overline{\mathfrak{A}} = \mathfrak{H}, \quad \mathfrak{N}' \subseteq \mathfrak{A}' \subseteq \overline{\mathfrak{A}}' = \mathfrak{H},$$
$$\overline{\mathfrak{W}}' = \overline{\mathfrak{B}} = \mathfrak{H} - \mathfrak{N}, \quad \overline{\mathfrak{W}} = \overline{\mathfrak{B}}' = \mathfrak{H} - \mathfrak{N}'.$$

Die Projektionsoperatoren von $\mathfrak{H} - \mathfrak{N}$, $\mathfrak{H} - \mathfrak{N}'$ seien bzw. E, E'.

Es gibt einen Operator W (überall definiert, stetig), der $\mathfrak{B}$ auf $\mathfrak{W}$, also $\mathfrak{H} - \mathfrak{N} = \overline{\mathfrak{B}}$ auf $\mathfrak{H} - \mathfrak{N}' = \overline{\mathfrak{W}}$ eineindeutig, linear, längentreu abbildet; W^* bildet ebenso $\mathfrak{B}'$ auf $\mathfrak{W}'$, also $\mathfrak{H} - \mathfrak{N}' = \overline{\mathfrak{B}}'$ auf $\mathfrak{H} - \mathfrak{N} = \overline{\mathfrak{W}}'$ ab; dabei ist bezüglich der zweitgenannten Abbildungen W^* zu W invers. W ist in $\mathfrak{N}$ Null, W^* ist in $\mathfrak{N}'$ Null. Es ist

$$W^*W = E, \quad WW^* = E'.$$

Es gelten die Beziehungen[20]:

$$\begin{aligned}A &= WB = CW, & A^* &= BW^* = W^*C, \\ B &= W^*A = A^*W, & C &= AW^* = WA^*,\end{aligned}$$

und aus ihnen folgt

$$C = WBW^*, \quad B = W^*CW.$$

Man beachte: B ist in $\mathfrak{N}$, C in $\mathfrak{N}'$ Null, also nur in $\mathfrak{H}-\mathfrak{N}$ bzw. $\mathfrak{H}-\mathfrak{N}'$ interessant. Hier aber gehen sie durch die längentreue Abbildung von $\mathfrak{H}-\mathfrak{N}$, $\mathfrak{H}-\mathfrak{N}'$ aufeinander, die W, W^* vermitteln, ineinander über. Dies ist die Verallgemeinerung des in Einl. 3, bzw. Anm. 11, erwähnten E. Schmidtschen Satzes.

Wenn 0 weder für A, noch für A^* (Punkt-) Eigenwert ist, d. h. wenn $Af = 0$ ebenso wie $A^*f = 0$ $f = 0$ zur Folge hat, so ist $\mathfrak{N} = \mathfrak{N}' = (0)$, $E = E' = 1$, also $W^*W = WW^* = 1$, d. h. W unitär, $W^* = W^{-1}$. In diesem Falle sind die Formeln von Satz 7 noch durchsichtiger.

II. Anwendungen.

1. Wenn $A^*A = AA^*$ ist, d. h. $B = C$, so folgt aus Satz 7: $A = WB = BW$, $A^* = BW^* = W^*B$, ferner $\mathfrak{N} = \mathfrak{N}'$, $E = F$, also $W^*W = WW^* = E$, also alle Operatoren B, W, W^* miteinander vertauschbar. Hieraus kann die Normalität von A (vgl. Einl. 3) auf verschiedene Weisen gefolgert werden, z. B. könnte man mit den Methoden von E., S. 115 (vor Satz 11*) eine simultane Eigenwertdarstellung für B, W, W^* und damit für A, A^* angeben. Wir schließen uns lieber an Definition in A., S. 116, Def. 6 an und schließen so. A, A^* haben dieselben Definitionsbereiche wie B, C, also beide denselben, sie bilden also im Sinne von A., S. 405, Def. 5 ein konjugiertes Paar. Mit der dortigen Bezeichnungsweise ist, wenn wir noch mit $E(\lambda)$ die zu B gehörige Zerlegung der Einheit bezeichnen,

$$(A)'' = (WB)'' \subseteq (W, B)'' = (W, W^*, \text{alle } E(\lambda))''.$$

Und diese Menge ist Abelsch, weil es die Menge der W, W^*, $E(\lambda)$ ist (vgl. A., S. 389 oben; W, W^* sind mit B vertauschbar, also auch mit allen $E(\lambda)$, A., S. 408, Satz 13).

Ist umgekehrt A normal, so haben die abgeschlossenen Operatoren A, A^* denselben Definitionsbereich (A., S. 416, Satz 21), und wenn A^*Af sowie AA^*g Sinn haben, so ist

$$(f, AA^*g) = (A^*f, A^*g)^{21} = (Af, Ag) = (A^*Af, g).$$

Wegen der Hypermaximalität von AA^* ist also AA^*f sinnvoll und gleich A^*Af. Also ist AA^* Fortsetzung von A^*A, da aber auch dieses hypermaximal ist, $A^*A = AA^*$. Also:

[21] Nach A., am soeben a. O., ist, wenn Ah Sinn hat, $\|Ah\| = \|A^*h\|$, also $(Ah, Ah) = (A^*h, A^*h)$. Hieraus schließt man, wenn Af, Ag Sinn haben, durch Betrachten von $\frac{f \pm g}{2}$, $\frac{f \pm ig}{2}$, daß auch $(Af, Ag) = (A^*f, A^*g)$ ist.

SATZ 8. *Sei A wie in Satz 2, 3, $B = \sqrt{A^*A}$, $C = \sqrt{AA^*}$. Dafür, daß A normal ist, ist $A^*A = AA^*$ charakteristisch oder auch $B = C$. Da A^*A, AA^* beide hypermaximal sind, genügt es sogar, wenn einer den anderen fortsetzt.*

Wenn A, A^* dieselben Definitionsbereiche haben und stets $\|Af\| = \|A^*f\|$ ist, so gilt dasselbe für B, C. $Bf \Leftrightarrow Cf$ bildet $\mathfrak{B}$ auf $\mathfrak{B}'$ eineindeutig, linear, längentreu ab, es kann daher ebenso auf $\overline{\mathfrak{B}} = \mathfrak{H} - \mathfrak{N}$, $\overline{\mathfrak{B}}' = \mathfrak{H} - \mathfrak{N}'$ ausgedehnt werden; aber es ist $\mathfrak{N} = \mathfrak{N}'$, $\overline{\mathfrak{B}} = \overline{\mathfrak{B}}'$. Dieser Operator heiße V_0, er bildet somit $\overline{\mathfrak{B}} = \overline{\mathfrak{B}}' = \mathfrak{H} - \mathfrak{N}$ auf sich ab, wir erweitern ihn zu einem linearen W_0, das in $\mathfrak{N}$ Null ist. Wie bei Satz 6 ist: $C = V_0 B$, also $C = W_0 B$ und

$$W_0^* W_0 = W_0 W_0^* = E.$$

Anwenden von $*$ ergibt $C = B W_0^*$, also führt W_0^* (innerhalb von $\mathfrak{H} - \mathfrak{N}$) den Definitionsbereich von C in denjenigen von B über, d. h. den von B in sich selbst. Dasselbe gilt für seine Inverse (in $\mathfrak{H} - \mathfrak{N}$) W_0. Bf und $Cf = W_0 Bf$ (beide sind in $\mathfrak{H} - \mathfrak{N}$) liegen also gleichzeitig im gemeinsamen Definitionsbereiche von B, C, d. h. B^2, C^2 haben denselben Definitionsbereich.

Da dort stets

$$(B^2 f, f) = (Bf, Bf) = \|Bf\|^2, \quad (C^2 f, f) = (Cf, Cf) = \|Cf\|^2,$$

also $(B^2 f, f) = (C^2 f, f)$ gilt, ist $B^2 = C^2$ — also $B = C$. Also:

SATZ 9. *Für die Normalität von A ist auch dieses charakteristisch: A, A^* haben denselben Definitionsbereich, und in ihm gilt stets $\|Af\| = \|A^*f\|$*[22].

2. Wir gehen zu den in Einl., 4. definierten Begriffen der metrischen Gleichheit und Fortsetzung von Operatoren über.

Aus Satz 6, sowie der beim Beweise von Satz 9 gemachten Überlegung, die zeigt, daß irgend zwei hermitesche, definite, hypermaximale Operatoren B, C, die metrisch gleich sind, auch wirklich gleich sein müssen (denn der Zusammenhang von B, C mit A, A^* wurde dort nicht benutzt), ergibt sich unmittelbar:

SATZ 10. *Jeder linear abgeschlossene $*$-Operator A ist einem und nur einem hermiteschen, definiten, hypermaximalen B metrisch gleich, d. i. das B von Satz 6: $B = \sqrt{A^*A}$.*

Jedes unstetige A hat echte metrische Fortsetzungen. Nach Satz 10 können wir A als hermitesch, definit, hypermaximal annehmen. Nach E., S. 108, Satz 49, ist A echte Fortsetzung eines linear abgeschlossenen

[22] In A., S. 423 oben, wurde gezeigt, daß es nicht hinreichend ist, wenn dies bloß auf irgend einer überall dichten Menge gilt.

hermiteschen Operators A', der sogar auf Kontinuum viele verschiedene Weisen wählbar ist, also ist A'^* (das $*$ ist, da A' $**$ ist) echte Fortsetzung von A. Also:

SATZ 11. *Sei A wie in Satz 10. Wenn er unstetig ist, so hat er echte metrische Fortsetzungen (sogar Kontinuum viele verschiedene).*

Wenn A stetig ist, ist sein Definitionsbereich abgeschlossen, und da er überall dicht sein muß, ist er dann gleich $\mathfrak{H}$. Hieraus, und aus Satz 11 folgt:

SATZ 12. *Sei A wie in Satz 10. A ist dann und nur dann stetig, wenn er überall sinnvoll ist.*

Dies ist die in Einl., 2. erwähnte Verallgemeinerung des Toeplitzschen Satzes[23], die übrigens auch direkt bewiesen werden kann.

Auch über die gleichzeitige metrische Fortsetzung mehrerer Operatoren, z. B. solcher, deren Definitionsbereiche keine gemeinsamen Punkte besitzen (vgl. M., S. 232, Satz 18), gelten sehr allgemeine Sätze, auf die der Verfasser demnächst einzugehen beabsichtigt.

[23] Math. Ann. 69 (1911), S. 321, vgl. auch E., S. 107–108, Satz 48, γ).

ALGEBRAS OF DEGREE 2^e AND PURE RIEMANN MATRICES.[1]

By A. Adrian Albert.

1. **Introduction.** The most recent discoveries in the theory of linear associative algebras have been due to the discussion of a new invariant number, the exponent of a normal division algebra. If A is any normal division algebra of degree n (order n^2) over F then the exponent of A is the least integer ϱ such that A^ϱ is a total matric algebra. H. Hasse has shown that the exponent of any *cyclic* algebra over an algebraic field Ω is equal to its degree. In the present paper I consider the corresponding question for *any* normal division algebra of degree a power of two over an algebraic field and prove that here again the exponent is equal to the degree. In particular I obtain the important theorem that *the only self-reciprocal normal division algebras over an algebraic field are algebras of degree one or two.*

The principal problem in the closely related theories of the complex multiplication of Riemann matrices, singular correspondences between algebraic curves, algebraic correspondences on an abelian variety, is that of a determination of the multiplication algebra D of any pure Riemann matrix. This algebra is a normal division algebra over an algebraic field. Either this field $\Omega = R(s)$ where R is the field of all rational numbers and s is a quantity the roots $\sigma_1, \cdots, \sigma_t$ of whose minimum equation are all real, or $\Omega = R(q, s)$ where s is as before and $q^2 = \mu(s)$ such that $\mu(\sigma_j) < 0$. I have called a Riemann matrix ω of the first or second kind according as $\Omega = R(s)$ or $\Omega = R(q, s)$. As an immediate corollary of my theorem on self-reciprocal algebras I show here that the multiplication algebra of any Riemann matrix of the first kind is either a field $R(s)$ or a generalized quaternion algebra over $R(s)$.

2. **The exponent of an algebra.** We shall use a large number of recently discovered theorems on normal division algebras, presenting them with references as to their proofs.

Theorem 1. *The exponent ϱ of any normal division algebra A of degree n is a divisor of n and every prime factor of n divides ϱ. Moreover A^α is a total matric algebra if and only if α is divisible by ϱ.*

Theorem 2. *Write any positive integer n in its unique form*

$$n = p_1^{e_1} p_2^{e_2} \cdots p_m^{e_m}$$

[1] Received October 24, 1931.

where the p_i are distinct primes. Every normal division algebra A of degree n over any non-modular field F is expressible as a direct product

$$A = B_1 \times B_2 \times \cdots \times B_m \tag{1}$$

of normal division algebras B_i of degrees $p_i^{e_i}$ respectively in one and only one way in the sense of equivalence and conversely.[2]

As an immediate consequence of the statement in Theorem 1 we have

COR. I. *If $n = p^e$, where p is a prime, then $\varrho = p^f$, $f \leqq e$.*

We also easily obtain

COR. II. *Let the exponents of the algebras B_i of* (1) *be ϱ_i respectively. Then $\varrho = \varrho_1 \cdot \varrho_2 \cdots \varrho_m$.*

For let $\alpha = \varrho_1 \cdot \varrho_2 \cdots \varrho_m = \varrho_i \alpha_i$. Evidently

$$A^\alpha = B_1^{\varrho_1 \alpha_1} \times B_2^{\varrho_2 \alpha_2} \times \cdots \times B_m^{\varrho_m \alpha_m}$$

is a direct product of total matric algebras and is a total matric algebra. Then $\alpha \geqq \varrho$ by Theorem 1. By the definition of ϱ the algebra $A^\varrho = B_1^\varrho \times B_2^\varrho \times \cdots \times B_m^\varrho$ is a total matric algebra. The algebras $(B_i)^\varrho$ are normal simple algebras whose orders are relatively prime and, by Theorem 2, must each be total matric algebras. Hence ϱ is divisible by each of the ϱ_i. Since the ϱ_i are relatively prime integers ϱ is divisible by their product α. Hence $\varrho \geqq \alpha$ and finally $\varrho = \alpha$.

We shall require my theorems

THEOREM 3. *Let y in A have grade s over F so that $n = st$, where n is the degree of A over F. If η is a scalar root of the minimum equation of y for F*

$$A \times F(\eta) = M \times B_0$$

where M is a total matric algebra of degree s over F and B_0 is a normal division algebra of degree t over its centrum $F(\eta)$. The algebra B of all the quantities of A which are commutative with y is a normal division algebra of degree t over $F(y)$ equivalent to B_0 as over $F(\eta)$.

THEOREM 4. *Let $n = p^e$, p a prime, and let x in A have grade n, the degree of A over F. Then there exists a field K of order r prime to n over F such that*

$$A' = A \times K$$

is a normal division algebra of degree n over K, and $K(x)$ of order n over K contains a sequence of sub-fields

$$K_e = K(x) > K_{e-1} > \cdots > K_0 = K$$

[2] Theorems 1 and 2 were first discovered by R. Brauer. An independent and comparatively short proof of these theorems will be found in my paper, *On direct products*, Trans. Amer. Math. Soc., vol. 33 (1931), pp. 690–711.

where K_i is a cyclic field of order p over K_{i-1}, K_i is a field of order p^i over K.

I have also proved[3]

THEOREM 5. *Let A be a normal division algebra of degree n over F and $F(\eta)$ an algebraic field of order r over F. Then*

$$A' = A \times F(\eta) = M \times B$$

where M is a total matric algebra of degree s over F and B is a normal division algebra of degree t over $F(\eta)$ such that

$$r = sk, \quad n = st$$

so that s divides r. There exists a total matric algebra K of degree k over F such that $A \times K$ contains a sub-algebra equivalent to B.

As an immediate corollary of Theorem 5 we may obtain a result which has been previously obtained by R. Brauer and E. Noether. It is, however, a very simple consequence of my more general considerations of Theorem 5.

COR. III. *Let $A \times F(\eta)$ be a total matric algebra so that $t = 1$. Then r is divisible by n.*

We also obviously have

COR. IV. *Let $r = s$. Then A contains a sub-algebra B_0 over a field $F(y)$ equivalent to B as over $F(\eta)$.*

At this point we shall introduce the notation for the general algebraic field over R, the field of all rational numbers. For various purposes it is better not to restrict our attention to *algebraic number fields* but to consider fields of the following type. Let $\varphi(\xi) = 0$ have degree h, rational coefficients and be irreducible in R. Let x be any quantity satisfying $\varphi(\xi) = 0$, so that x may be a matrix, an algebraic number, or simply a quantity in an algebra. Then the set $\Omega = R(x)$ of all rational functions of x with rational coefficients is a field called an algebraic field over R. Evidently $R(x)$ is simply isomorphic with $R(\eta)$ where η is any ordinary complex root of $\varphi(\xi) = 0$, that is η an algebraic number. We shall usually use the notation Ω for $R(x)$ in the remainder of this paper.

I have proved[4]

THEOREM 6. *The exponent of any normal division algebra of degree four over Ω is four.*

We have now the necessary tools for our further new considerations.

[3] For Theorems 4 and 5 see Theorems 23, 14 and 18 of my paper *On direct products* (loc. cit.). For Theorem 3 see my paper *Algebras over an algebraic field* to be published probably in the October 1931 number of the Bull. Amer. Math. Soc.

[4] In a paper entitled *Normal division algebras of order sixteen over an algebraic field* which has been accepted for publication by the Trans. Amer. Math. Soc.

3. **Algebras of degree 2^e.** Let A be any normal division algebra of degree n (order n^2) over F, any non-modular field. As is well known, for every integer $\alpha > 1$

$$A^\alpha = M^{\alpha-1} \times M_\alpha \times A_\alpha \tag{2}$$

where M is a total matric algebra of degree n, M_α is a total matric algebra of degree s_α, and A_α is a normal division algebra of degree t_α over F such that

$$s_\alpha t_\alpha = n. \tag{3}$$

In particular let $n = p^e$, p a prime, and let $\alpha = p$. Then

$$A^p = M^{p-1} \times M_p \times A_p. \tag{4}$$

We shall prove

THEOREM 7. *The degree t_p of A_p is a divisor of p^{e-1}.*

For by Theorem 4 there exists an algebraic field K of order r prime to p over F such that

$$C = A \times K$$

is a normal division algebra over K and contains a field K_{e-1} of order p^{e-1} over K. Evidently

$$C^p = M^{p-1} \times M_p \times C_p$$

where $C_p = A_p \times K$ is a normal division algebra over K by Theorem 5. Let $K_{e-1} = K(y)$ so that y, which generates K_{e-1} has $\varphi(\xi) = 0$ of degree p^{e-1} as its minimum equation with respect to K. Let η be a scalar root of $\varphi(\xi) = 0$ so that, by Theorem 3, if $C' = C \times K(\eta)$ then

$$C' = M_0 \times B_0$$

where M_0 is a total matric algebra and B_0 has degree p. But B_0^p is a total matric algebra by Theorem 1. Hence so is $(C')^p$. Also now

$$(C')^p = M^{p-1} \times M_p \times (C_p)'$$

is a total matric algebra where $(C_p)' = C_p \times K(\eta)$. By Cor. III the degree of C_p is a divisor p^{e-1} since $(C_p)'$ is a total matric algebra. The degree of C_p is evidently the degree t_p of A_p so that t_p divides p^{e-1}.

We shall apply Theorem 7 to the case where $F = \Omega$, an algebraic field over R, and with $p = 2$. We shall demonstrate

THEOREM 8. *Let A be a normal division algebra of degree $n > 1$ over Ω. Then A^2 is a total matric algebra if and only if $n = 2$.*

For if $n = 2$ then A^2 is a total matric algebra by Theorem 1. Conversely let A^2 be a total matric algebra. Since $n > 1$, $\varrho > 1$. Hence ϱ

has its least value 2. By Theorem 1 the degree n of A is a power 2^e of 2. Let $e \geqq 2$. By Theorem 4 there exists an algebraic field K of odd order over Ω such that $C = A \times K$ is a normal division algebra of degree 2^e over K and contains a sub-field $K(y)$ of degree 2^{e-2}, a positive or zero power of two. Let η be a scalar root of the minimum equation of y for K so that $K(\eta)$ is simply isomorphic with $K(y)$. By Theorem 3

$$C \times K(\eta) = G \times B$$

where G is a total matric algebra and B is a normal division algebra of degree four over $K(\eta)$. But A^2 is a total matric algebra so that so is C^2. Hence $G^2 \times B^2$ is a total matric algebra and so is B^2. By Theorem 6 the exponent of B is four, a contradiction.

Hence e is at most unity. But n is greater than unity so that e is not zero. Hence $e = 1$, $n = 2$.

We note here that Theorem 8 and its proof would be valid when we replace two by any prime p provided we could prove Theorem 6 for four replaced by p^2. I have, as yet, been unable to complete the proof of this latter theorem.

We shall now pass to an important consequence of Theorem 8. R. Brauer has proved[5]

LEMMA. *Let A be a normal division algebra over F and let A_0 be reciprocal to A. Then $A \times A_0$ is a total matric algebra.*

As a consequence of this lemma and Theorem 8 we have

THEOREM 9. *The only self-reciprocal normal division over an algebraic field Ω are algebras of degree unity or two.*

For if A is self-reciprocal then $A^2 = A \times A_0$ is a total matric algebra. Hence either the degree of A is unity or, by Theorem 8, the degree of A is two.

4. The exponent of any normal division algebra of degree 2^e. We shall now obtain the principal result of this paper. Let A be a normal division algebra of degree 2^e over an algebraic field Ω and let

$$A^2 = M \times H \times B$$

where M is a total matric algebra of degree 2^e, H is a total matric algebra of degree 2^g, and B is a normal division algebra of degree 2^f such that $f + g = e$. By Theorem 7 the integer f is at most $e - 1$ and, by Theorem 8, at least unity. As is well known there exists an algebraic field Δ of order 2^f over Ω such that $B \times \Delta$ is a total matric algebra. Also $A' = A \times \Delta = G \times C$ where G is a total matric algebra of degree s

[5] Math. Z., vol. 30 (1929), pp. 79–107.

and C is a normal division algebra of degree t over $\mathcal{A}$ such that $st = 2^e$ and s divides 2^f, the order of $\mathcal{A}$ by Theorem 5. Assume at this point that $f < e-1$ so that $t \geqq 4$. Since $A^2 = M \times H \times B$ we have $(A')^2 = G^2 \times C^2 = M \times H \times B'$ where M, H, and $B' = B \times \mathcal{A}$ are total matric algebras. Hence C^2 is a total matric algebra. The degree of C is at least four and, by Theorem 8, C^2 is not a total matric algebra. This contradicts our hypothesis that $f < e-1$. Hence $f = e-1$ and we have

LEMMA. *Let A be a normal division algebra of degree 2^e over Ω. Then $A^2 = M_2 \times A_2$ where M_2 is a total matric algebra and A_2 is a normal division algebra of degree 2^{e-1}.*

Assume as the basis of an induction on e that the exponent of any normal division algebra of degree $2^f < 2^e$ is 2^f, a result true for the degree two by Theorem 1. Let then A have degree 2^e so that the exponent of A is 2^m, a divisor of 2^e. Thus $m \leqq e$. Write $\varrho = 2\sigma$. The direct product $A^2 = M_2 \times A_2$ where A_2 is a normal division algebra of degree 2^{e-1}, and by the hypothesis of our induction has exponent 2^{e-1}. Also $A^\varrho = A^{2\sigma} = (M_2)^\sigma \times (A_2)^\sigma$ is a total matric algebra so that so is $(A_2)^\sigma$ and, by Theorem 1, $\sigma \geqq 2^{e-1}$. But then $\varrho = 2\sigma \geqq 2^e$. Hence $\varrho = 2^e$ and our induction is complete. We have proved

THEOREM 10. *The exponent of any normal division algebra of degree 2^e over an algebraic field Ω is 2^e.*

We note here that in the above theorem and proof we would again have the result true for any prime p providing we could prove the corresponding Theorem 6.

Let $\varphi(\xi) = 0$ have coefficients in Ω and be irreducible in Ω. We say that $\varphi(\xi) = 0$ belongs to a normal division algebra A over Ω if there exists a quantity y in A with $\varphi(\xi) = 0$ as its minimum equation. As a simple generalization of Theorem 10 we have

THEOREM 11. *Let A be a normal division algebra of degree 2^e over an algebraic field Ω and let $0 \leqq t < e$, $s = 2^{e-t}$. Then*

$$A^s = M_s \times A_s$$

where M_s is a total matric algebra over Ω and A_s is a normal division algebra over Ω with degree 2^t and which is such that every equation of degree $2^r \leqq 2^t$ which belongs to A_s belongs to A and conversely.

For since the exponent of A is 2^e and $s = 2^{e-t}$ the exponent of A_s, the normal division algebra component of A^s, is 2^t. By Theorem 10 the degree of A_s is 2^t. Let $\varphi(\xi) = 0$ have degree 2^r and belong to A_s so that the order of the algebraic field $\Omega(\eta)$ defined by this equation is $2^r \leqq 2^t$. By Theorem 3 $A_s \times \Omega(\eta) = H \times B$ where H is a total matric algebra and B is a normal division algebra of degree 2^{t-r} over $\Omega(\eta)$. By

Theorem 5 we have $A' = A \times \Omega(\eta) = M \times C$ where M is a total matric algebra and C is a normal division algebra of degree 2^c, a divisor of 2^e. But $(M \times C)^s = (A')^s$ over $\Omega(\eta)$ has B as its division algebra component. The exponent of B is 2^{t-r} so that the exponent of C^s is 2^{t-r}. But then the exponent of C is $2^c = (2^{t-r})(2^{e-t})$ since $s = 2^{e-t}$. Hence $c = e-r$. By Corollary IV algebra A contains a sub-algebra C_0 over $\Omega(y)$ equivalent to C as over $\Omega(\eta)$ and $\varphi(\xi) = 0$ is the minimum equation of y for Ω and belongs to A.

Conversely let $\varphi(\xi) = 0$ have degree $2^r \leqq 2^t$ and belong to A. With $\Omega(\eta)$ as before we have $A' = A \times \Omega(\eta) = M \times C$ as before. Also $(A')^s = Q \times (A_s)'$ where Q is a total matric algebra and $(A')_s = A_s \times \Omega(\eta) = H \times B$ as before. The degree of C is 2^{e-r} by Theorem 3. The degree of B, the division algebra component of C^s is $2^{e-r} \cdot s^{-1} = 2^{t-r}$ by Theorem 10. Applying Corollary IV as previously we have shown that $\varphi(\xi) = 0$ belongs to A_s.

5. **Application to the theory of Riemann matrices.** A matrix ω with p rows and $2p$ columns of complex elements is called a Riemann matrix if there exists a rational alternate $2p$-rowed square matrix C such that $\omega C \omega' = 0$ and $\sqrt{-1}\, \omega C \bar{\omega}'$ is a positive definite Hermitian matrix. It is well known that C and the matrix

$$\left\| \begin{matrix} \omega \\ \bar{\omega} \end{matrix} \right\|$$

are both non-singular. A $2p$-rowed rational square matrix A is said to define a multiplication a of ω if

$$a\omega = \omega A$$

for some complex p-rowed square matrix a. The set of all multiplications of a Riemann matrix ω is a linear associative algebra over the field R of all rational numbers, and when ω is pure this algebra D is a division algebra. If A defines a multiplication of ω then so does

$$A^* = CA'C^{-1}.$$

This implies the existence of an operation in D such that

$$(a+b)^* = a^* + b^*, \quad (\lambda a)^* = \lambda a^*, \quad (a^*)^* = a$$

for every rational λ and a and b in D, while

$$(ab)^* = b^* \cdot a^*, \qquad (a, b \text{ in } D).$$

A multiplication a is called symmetric if $a^* = a$ and skew-symmetric if $a^* = -a$. Algebra D is a normal division algebra over its centrum $R(q)$ of order ε. Let the degree of D over $R(q)$ be n. As is known[6] either $q = s = s^*$ a symmetric element of algebra D or $R(q) = R(u, s)$ where s is symmetric and u is skew-symmetric such that u^2 is in $R(s)$. In both cases the field $R(q)$ has the properties as given in the introduction to this paper and we call ω a pure Riemann matrix of the first or second kind in the respective cases. The theory of matrices of the first kind has seemed the more complicated theory up to the present.[7] We shall now make a complete determination of D for matrices of the first kind and show how simple this case really is.

Let D be an algebra of the first kind. Then D is a normal division algebra over $R(s)$ where $s = s^*$. Let us set up the correspondence $a^* \simeq a$ for every a of D. This is a self-correspondence of D in which every element of the centrum of D is self-corresponding,

$$(a+b)^* = a^* + b^* \simeq a + b, \quad (\lambda a)^* = \lambda a^* \simeq \lambda a, \quad (ab)^* = b^* a^* \simeq ab,$$

for every a and b of D and λ of $R(s)$. Hence D is a self-reciprocal algebra over its centrum $R(s)$, an algebraic field over R. By Theorem 9 the degree n of D is unity or two, and D is either a field or a generalized quaternion algebra.

THEOREM 12. *Let the multiplication algebra D of a pure Riemann matrix ω be an algebra of the first kind. Then D is either a field of symmetric quantities or a generalized quaternion algebra over such a field.*

---

[6] For a study of these and related results see my paper in the Palermo Rendiconti, vol. 55 (1931), pp. 57–115.

[7] See for example Rosati's extremely complicated discussion of the singularity index of such an algebra. In his last paper *Sulle matrici di Riemann,* Rendiconti del Circolo Matematico di Palermo, vol. 53 (1929), pp. 79–134, he proved that the singularity index k of D had the values t, the order of $R(s)$ when $n = 1$, and in all other cases had the two possible values $(1/2)\,tn(n+1)$ or $(1/2)\,n(n-1)\,t$ according as D does or does not contain a symmetric quantity of grade n with respect to the centrum $R(s)$. For $n = 2$ these values become $3t$ and t. It is almost trivial to prove that these values are correct for what I have now shown to be *the only possible cases.* The cases are now distinguished by the existence or non-existence of any symmetric quantity in D not in $R(s)$.

UNIVERSITY OF CHICAGO.

A FEW DESIRABLE MODIFICATIONS IN THE LITERATURE OF GROUP THEORY.[1]

By G. A. Miller.

Among the numerous errors that appear in the literature of group theory no one seems more striking to the present writer than the one which is found in H. Weber's *Lehrbuch der Algebra*, vol. 2 (1896), p. 54, where it is stated that the most important example of a commutative group is furnished by the natural numbers when they are combined by ordinary multiplication. According to the definition of a group given earlier in the same volume these numbers do not constitute a group at all when they are thus combined since it is not always possible to find such a number which satisfies the equation $a\,x = b$, when a and b are given natural numbers. This error was corrected by H. Weber himself in the second edition of this volume, 1899, p. 4, but in correcting it the author failed to refer explicitly to the place where it appeared and hence the reader is apt to overlook the fact that it is an actual correction.

This error will appear especially striking when it is remembered that its author seems to have been the first to formulate explicitly satisfactory sets of postulates for a finite abstract group (1882) and also for an infinite abstract group (1893). It should, however, be observed that A. Cayley used such a set of postulates for a finite abstract group at a much earlier date (1854), but he did not formulate all the postulates of the set as clearly and definitely as H. Weber did twenty-eight years later. These historic statements may be of less interest to some readers thereof than the consideration of a slight modification in the definition of an abstract group which would enable us to include under this term the closed system formed by the natural numbers when they are combined by ordinary multiplication, especially since the effects of such modifications in sets of group postulates have thus far received but little attention, and this set of numbers is of such great historic interest in the development of mathematics.

One of the well known sets of postulates of an abstract group whose order may be either finite or infinite is that in addition to the associative law the distinct elements of the group must be such that if any two of the three symbols x, y, z, in the equation

$$x\,y = z$$

[1] Received August 1, 1930.

are replaced by the same element or by two distinct elements of the group the third of these symbols can be replaced by one and only one such element. If we modify this statement by saying that the symbols x and y satisfy this condition, and wherever either of the other two pairs of symbols can be replaced in this equation by elements of the group then this condition must also be satisfied with regard to this pair, we clearly have a set of postulates for a group of finite or infinite order which is equivalent to the commonly used set noted above when the group is of finite order and applies also to the totality of the natural numbers when these are combined by ordinary multiplication, and hence classes this system with the infinite groups.

From what precedes it results that the closed system which H. Weber called the most important example of a commutative group in the first edition of his noted *Lehrbuch der Algebra* he regarded as no group at all in the second edition thereof, but that he might have continued to regard it as a group if he had made a slight change in his postulates without affecting their use as regards abstract groups of finite order. He might also have replaced the given example in the second edition by the group formed by the rational numbers, not including zero, when these numbers are combined by ordinary multiplication, without making any change in his group postulates. What he actually did was to replace the given example of a supposed group of infinite order according to his postulates by an example of the group formed by certain integers when they are combined by ordinary multiplication with respect to a given modulus. This group he calls here the most important example of a finite commutative group.

In view of the great influence of the work under consideration these modifications and judgments are naturally of unusual interest even if such a broad statement as "most important example" should be regarded as resulting from an effort to arouse interest in the group under consideration rather than as an expression of finality. For instance, it would be impossible to disprove the statement that the group formed by the positive rational numbers when they are combined by ordinary multiplication is the most important example of a commutative group. A similar claim might be made for the group formed by the totality of integers, including zero, when they are combined by ordinary addition. Such claims are, however, of questionable propriety in scientific work unless they are properly modified. This is especially true when they relate to historical considerations. A necessary and sufficient condition that a set of natural numbers constitutes a group when combined by multiplication with respect to a given modulus was developed in these *Annals*, vol. 18 (1905), p. 44.

One of the most unsatisfactory but widely used definitions in the literature of the theory of groups relates to multiply transitive substitution groups. It is frequently stated that a group of degree n is k times transitive if every set of k of these n letters is transformed under the group into every other such set. In particular, it is often said that a substitution group is doubly transitive if every pair of its n letters is transformed under the group into every possible pair of these letters. To exhibit the fact that it is necessary to modify this definition it may be noted here that the group of degree 7 and of order 21 transforms every one of its 21 pairs of letters into every such pair and yet it is only simply transitive since the subgroup composed of all its substitutions which omit one letter is intransitive. The general definition should specify that every ordered set of k letters is transformed into every such set, but this is frequently not done. Cf. Pascal's *Repertorium der höheren Mathematik*, vol. 1 (1910), p. 211.

It is obvious that when every combination of k letters is transformed under the group into every such combination then this must also be true as regards the combinations of $n-k$ letters. Hence the definition of multiply transitive groups which is widely used would seem to imply that every transitive group of degree n is also $n-1$ times transitive. It is, however, of greater interest to note here that whenever every combination of k letters of a transitive group is transformed under the group into every such combination then the group must be primitive unless k is either 1 or $n-1$. This results directly from the fact that if such a group were imprimitive it would be possible to select two combinations of k letters in such a manner that the letters would be divided differently among its systems of imprimitivity. Such combinations could not be transformed into each other under the group and hence it results that *if every non-ordered set of k letters of a substitution group of degree n is transformed under the group into every such set then the group must be primitive except possibly when k is either* 1 *or* $n-1$.

It may be noted in this connection that a necessary and sufficient condition that a substitution group of order g and of degree n is at least r times transitive is that the order of every subgroup composed of all the substitutions which omit any given set of r letters is $g/n(n-1)\cdots(n-r+1)$, and that the order of every such subgroup is greater than this quotient whenever this group is less than r times transitive. Hence we can determine the degree of transitivity of any substitution group of degree n as follows: Consider successively subgroups composed of all the substitutions of such a group which omit $1, 2, \cdots, r$ letters until we arrive at such a group whose order exceeds the quotient obtained by dividing g by $n(n-1)\cdots(n-r)$. If r is the first number for which this happens then the group is exactly

r times transitive and vice versa. In particular, the group is intransitive when this happens for $r = 0$, and vice versa.

While the definitions of imprimitive substitution groups are usually more satisfactory than those relating to multiply transitive groups some of the former definitions are still worded in such a way as to convey the notion to many readers that an imprimitive substitution group is necessarily a composite group. In fact, E. Netto explicitly stated this as a theorem in his well known *Substitutionentheorie*, 1882, p. 94, which is still widely used. This error was, however, corrected in the English translation of this book, 1892, p. 106; but even in such an excellent work as H. Weber's *Kleines Lehrbuch der Algebra*, 1912, there may be found, p. 213, a definition of an imprimitive group which seems to imply that such a group is necessarily composite. It is true that at the bottom of the same page there appears also a second and satisfactory definition of this term, but this is called an alternative definition and hence it fails to remove the objectionable feature of the other as regards the point in question.

In view of the great usefulness of this work and the valuable chapter on the theory of groups contained therein, it may be desirable to refer here also to a widespread historical error which appears therein as a footnote, p. 196, and credits R. Dedekind (1894) with priority as regards the use of double co-sets. On the contrary, A. L. Cauchy made considerable use of double co-sets as early as 1844 in his *Exercices d'analyse*, vol. 3, p. 245. Not only did he prove here some fundamental theorems as regards the number of distinct operators found in certain double co-sets with respect to substitution groups but he also used these co-sets to establish the following fundamental theorem, known by his name: If the order of a substitution group is divisible by a prime number then this group must involve a substitution whose order is equal to this prime. As is well known Sylow's theorem can easily be proved by means of this theorem.

It is perhaps still more surprising that H. Weber overlooked at the place noted above the valuable developments relating to double co-sets published by G. Frobenius in *Crelle's Journal*, vol. 101, p. 273, since this article by Frobenius appeared seven years prior to the work of Dedekind cited by him. In the same chapter, p. 192, H. Weber gives credit to O. Hölder for the symbol P/N, now commonly used to represent the quotient group of the group P with respect to its invariant subgroup N. On the contrary, this symbol was used much earlier by C. Jordan for the same purpose. It is not implied here that such slight modifications detract much from the usefulness of the chapter in question, but the fact that the former of these errors is repeated, in substance, in the article on

groups published in the fourteenth edition of the *Encyclopaedia Britannica*, vol. 10, p. 913, is evidence of the desirability of the suggested modifications.

Errors appearing in works of reference which are frequently consulted are especially apt to be a source of trouble for the student. It may therefore be desirable to include in this list an error which is found in the article on finite discontinuous groups published in the *Encyclopédie des Sciences Mathématiques*, t. 1, vol. 1, p. 584, but which was not included in the various lists of corrections published in the "Tribune publique", issued in connection with this encyclopedia. The error in question relates to the division of the operators of the group of isomorphisms of a given group into classes with respect to the invariant subgroup composed of the inner isomorphisms of this group. Two such operators are here substantially said to belong to the same class if they correspond to inverse operators in the quotient group with respect to this invariant subgroup, but it is stated incorrectly here that all these classes contain the same number of isomorphisms. As a matter of fact it is evident that the class composed of the inner isomorphisms, called the identity class, contains the same number of operators as each of the classes which correspond to operators of order 2 in the given quotient group, while the number of the operators in each of the other classes, when such classes exist, is always just twice as large.

The preceding suggestions as regards desirable modifications in the literature relating to group theory do not aim to be complete but they aim to bring out a few points of interest and certain extensions to the known facts relating to these elementary points. Probably the most extensive list of errors committed by renowned mathematicians appears in various volumes of the *Intermédiaire des Mathématiciens* in answer to a request made by E. Maillet for instances of such errors. This request appeared as question number 2855, vol. 11, (1904), p. 285, of this periodical and the replies thereto extend through a number of years but they contain very little relating to the theory of groups. The above list may serve to supply to some extent similar material for those interested in this particular field as well as to exhibit a few points of theoretic interest relating thereto.

THE LINEAR FORM OF NUMBERS REPRESENTED BY A HOMOGENEOUS POLYNOMIAL IN ANY NUMBER OF VARIABLES.[1]

By Morgan Ward.

1. In this paper I obtain the following necessary conditions that all of the numbers properly represented by a homogeneous polynomial in any number of variables may be of one or the other[2] of the linear forms

$$nz,\ nz+a_1,\ \cdots,\ nz+a_r. \tag{1}$$

n here is any integer, and $a_1, \cdots, a_r$ are r distinct integers less than n and prime to it.[3]

THEOREM 1. *If all of the numbers properly represented by the homogeneous polynomial of degree N*

$$H = H(x_1, x_2, \cdots, x_k) = \sum_{(s)} h_{(s)} x_1^{s_1} \cdot x_2^{s_2} \cdots x_k^{s_k} \tag{2}$$

(all the $h_{(s)}$ integers)

are of one or the other of the forms (1), *and if*

$$n = p_1^{b_1} \cdots p_L^{b_L}$$

is the resolution of n into its prime factors, then it is necessary that the least common multiple of the numbers

$$p_1^{b_1-1}(p_1-1),\ \cdots,\ p_L^{b_L-1}(p_L-1)$$

divide rN.

We shall denote this least common multiple by $\lambda(n)$.

THEOREM 2. *Under the hypotheses of Theorem* 1, *the r numbers $a_1, \cdots, a_r$ in* (1) *must form q complete co-sets of the group G_τ of the N^{th} powers of the elements in the totient group of n.*

---

[1] Received January 21 and April 13, 1931.

[2] The form nz may be omitted without invalidating the theorems. Each of the other forms is assumed actually to occur.

[3] The problem becomes rather unwieldy if we remove the restriction that the a_i be prime to n. The simplest case is when all the numbers representable by the form are divisible by n; for polynomials in two variables, we have essentially the problem of determining all residual polynomials modulo n. See Dickson, Introduction to the Theory of Numbers, Chicago, (1929), Chapter II.

From Theorem 1, we see that[4] $\lambda(n) \leqq rN$. Since $\lambda(n)$ tends to infinity with n, we have the following corollary:

COROLLARY. *For a given r and N in* (1) *and* (2), *there are only a finite number of values of n satisfying the hypotheses of Theorem* 1.

Furthermore, we see from Theorem 2 that we must have

$$r = q\tau, \tag{3}$$

where τ is the order of the group G_τ.

2. Theorem 2 is readily established as follows.

Assume that the hypotheses of Theorem 1 are satisfied. Then for each a_i we can find a set of co-prime integers $c_1, c_2, \cdots, c_k$ such that[5]

$$H(c_1, c_2, \cdots, c_k) \equiv a_i \bmod n, \qquad (a_i, n) = 1.$$

Let s denote any integer prime to n. Then it is possible to choose k integers $z_1, z_2, \cdots, z_k$ so that the numbers

$$d_1 = z_1 n + s c_1, \; d_2 = z_2 n + s c_2, \; \cdots, \; d_k = z_k n + s c_k$$

are co-prime. The number $m = H(d_1, d_2, \cdots, d_k)$ is accordingly properly represented by the form H. Hence since H is homogeneous of degree N,

$$m \equiv s^N H(c_1, c_2, \cdots, c_k) \equiv s^N a_i \not\equiv 0 \bmod N.$$

But m must be congruent modulo n to some one of the a_i; therefore the numbers

$$s^N a_i, \qquad (i = 1, 2, \cdots, r), \quad (s, n) = 1, \tag{4}$$

are all congruent modulo n to one or the other of the numbers a_i in (1).

Now if $G_{\varphi(n)}$ denotes the totient group of n, the Nth powers of all the elements of $G_{\varphi(n)}$ form a sub-group G_τ of order $\tau \leqq \varphi(n)$. Hence for a given a_i, the numbers (4) are congruent modulo n to the τ numbers of some co-set of G_τ in $G_{\varphi(n)}$. Since the numbers a_i are all distinct and all in $G_{\varphi(n)}$, Theorem 2 follows.

The proof of Theorem 1 is now immediate. For if g is any element of G_τ, $g^\tau \equiv 1 \bmod n$. Hence since the elements of G_τ are congruent to the Nth powers of the elements of $G_{\varphi(n)}$, $s^{N\tau} \equiv 1 \bmod n$, so that by (3)

[4] That rN is actually the "best possible" maximum for $\lambda(n)$ is shown by the case $N = 3$, $r = 2$, $n = 7$, $a_1 = 1$, $a_2 = 6$. For $H = (x_1 + x_2 + \cdots + x_k)^3$, we find that $\lambda(n) = rN = 6$.

[5] We use when convenient the standard notation $(a, b, \cdots, c)$ for the greatest common divisor of the numbers $a, b, \cdots, c$.

$$s^{\tau N} \equiv 1 \mod n$$

for every integer s prime to n.

Accordingly, if $\lambda(n)$ is the least positive value of u such that $s^u \equiv 1 \mod n$ for all integers s prime to n, $\lambda(n)$ divides τN.

But if $n = p_1^{b_1} \cdots p_L^{b_L}$ is the resolution of n into its prime factors, $\lambda(n)$ is precisely[6] the L. C. M. of $p_1^{b_1-1}(p_1-1), \cdots, p_L^{b_L-1}(p_L-1)$.

3. As an application of these theorems, let us consider the problem[7] of obtaining primitive binary forms

$$H(x_1; x_2) = \sum_{s=0}^{N} h_s x_1^{N-s} x_2^s$$

of degree N such that the prime factors of all of the numbers properly represented by H are either divisors of n or of the form $nz \pm 1$ (so that n is necessarily even). Examples of such forms, due to Lehmer, place cited, are $x^3 + 16x^2y - 51xy^2 - y^3$ for $n = 14$ and $x^3 - 18x^2y + 69xy^2 - y^3$ for $n = 18$.

These forms are obviously included in the more general category of forms which properly represent only numbers of the types nz, $nz + a_1$, $nz + a_2$ with a_1, a_2 prime to n. Hence by Theorem 1 and equation (3), we must have $\lambda(n)$ a divisor of $2N$ and $\tau \leqq 2$.

For example, suppose that $N = 3$. $\lambda(n) = 1$ has the solution $n = 2$; $\lambda(n) = 2$ the four solutions 3, 4, 6, 12 while $\lambda(n) = 6$ has the twelve solutions 7, 9, 14, 18, 21, 28, 36, 42, 63, 84, 126, 252. Of the even values of n, the cases $n = 2$, 4 and 6 are trivial since every prime save 2 is of the form $2k+1$ and $4k \pm 1$ and every prime save 2 and 3 is of the form $6k \pm 1$. On the other hand in the cases 12, 28, 36, 42, 84, 126 and 252, $\tau > 2$. Hence $n = 14$ and $n = 18$ are the only non-trivial values of n for which such cubic forms can exist. In a similar manner one can show that 22 is the only non-trivial value of n for which such quintic forms can exist, and that there are no non-trivial values of n for septimic forms.

[6] Dickson, Work cited, p. 19.

[7] See D. H. Lehmer, An Extended Theory of Lucas' Functions. Annals, Vol. 31 (1930), p. 436. Dr. Lehmer has informed me that since the paper was written he has considerably extended his results on this problem.

A PROOF OF A CERTAIN CONGRUENCE.[1]

By W. L. G. Williams.

1. Frank Morley gave a proof of the congruence

$$(p-1)! \equiv (-1)^{(p-1)/2}\, 2^{2(p-1)} \left(\frac{p-1}{2}!\right)^2, \mod p^3$$

in these Annals, vol. IX (1895), pp. 168–170. We present here a proof of an entirely different nature. In the congruence as it appears above p is a prime number greater than 3; for later use we show also that the congruence (A) holds, mod n^2, if n is any odd number.

For any odd integer n, greater than unity, from the obvious congruence

$$n-2i \equiv -[n-2(n-i)], \mod n,$$

we see that

$$(n-2)(n-4)\cdots[n-2(n-1)] \equiv I n^2 + 2^{n-1}(n-1)!, \mod n^3,$$

where I is an integral multiple of

$$(n-1)!\left[\frac{1}{1\cdot 2}+\frac{1}{1\cdot 3}+\cdots+\frac{1}{(n-2)(n-1)}\right].$$

If n is an odd prime >3, by a well known theorem[2], $I \equiv 0$, mod n. Thus, the following three congruences hold modulis n^2 and n^3, according as the odd number n is not or is a prime >3:

$$(n-2)(n-4)\cdots[n-2(n-1)] \equiv 2^{n-1}(n-1)!,$$

$$[(n-2)(n-4)\cdots 1]^2 (-1)^{(n-1)/2} \equiv 2^{n-1}(n-1)!,$$

(A) $$[(n-1)!]^2 \equiv (-1)^{(n-1)/2}\, 2^{n-1}\,[2\cdot 4\cdot 6\cdots(n-1)]^2 (n-1)!,$$

whence

$$(n-1)! \equiv (-1)^{(n-1)/2}\, 2^{2(n-1)} \left(\frac{n-1!}{2}\right)^2, \mod n^3, \quad n \text{ prime}.$$

This completes the proof of the results stated.

2. There is a simple generalization of the result of Section 1, which we obtain by considering

(1) $$a_1, a_2, \cdots, a_{2m}$$

the $\varphi(n)$ numbers less than and prime to the odd integer n. As above,

[1] Received August 28, 1931.

[2] Cf., for example, N. Nielsen, Traité élémentaire des nombres de Bernoulli, p. 327.

$$(n-2a_1)(n-2a_2)\cdots(n-2a_m) \equiv 2^{2m}\, a_1 a_2 \cdots a_{2m}, \mod n^2,$$

$$(2) \qquad (n-2a_1)(n-2a_2)\cdots(n-1) \equiv (-1)^m\, 2^{2m}\, a_1 a_2 \cdots a_{2m}, \mod n^2.$$

If we designate by $\pi(e)$ and $\pi(o)$, respectively, the products of all the even and all the odd numbers of the set (1). Then (2) may be written

$$(\pi(o))^2 \equiv (-1)^m\, 2^{2m}\, \pi(e)\, \pi(o), \mod n^2,$$

whence

$$\pi(o) \equiv (-1)^{\frac{1}{2}\varphi(n)}\, 2^{\varphi(n)}\, \pi(e), \mod n^2.$$

McGILL UNIVERSITY,
August 10, 1931.

A SET OF POSTULATES FOR PLANE GEOMETRY, BASED ON SCALE AND PROTRACTOR.[1]

By George D. Birkhoff.

Introduction. Some years ago, in attempting to present the simplest geometric facts in popular form, I realized how plane geometry might be approached readily via the facts embodied in the scale and protractor.[2]

The purpose of the present article is to present the corresponding set of postulates in rigorous mathematical form. From the purely mathematical point of view these postulates are hardly to be classed with other sets among which, in this country alone, may be recalled those of Veblen, Veblen and Young, Schweitzer and Huntington. On account of its immediacy and possible usefulness, however, the new set possesses an obvious interest.

1. **Undefined elements and relations.** The undefined elements are (a) *points*, designated by $A, B, \cdots$, and (b) certain classes of points called (straight) *lines*, designated by $l, m, \cdots$.

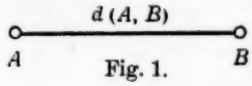

Fig. 1.

The undefined relations are (c) *distance between any two points* A, B, designated by $d(A, B)$, a real non-negative number with $d(A, B) = d(B, A)$ and (d) *angle formed by three ordered points* A, O, B $(A \neq O, B \neq O)$[3] designated by $\angle AOB$, a real number (mod 2π).

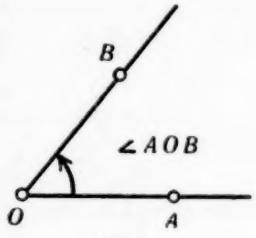

Fig. 2.

The point O is called the vertex of the angle.

2. **The postulate of line measure.** The postulate of line measure is taken as follows:

Postulate I. *The points* $A, B, \cdots$ *of any line* l *can be put into* (1, 1) *correspondence with the real numbers* x *so that* $|x_B - x_A| = d(A, B)$ *for all points* A, B.

Evidently the scale or marked ruler embodies this postulate.

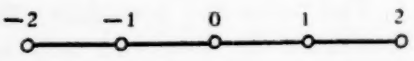

Fig. 3.

[1] Received August 8, 1931.

[2] *The Origin, Nature and Influence of Relativity*, Macmillan (1926), Chapter II, The Nature of Space and Time. See also a paper, written by my colleague Professor Ralph Beatley and myself *A New Approach to Elementary Geometry*, Yearbook of the National Association of Mathematics Teachers, 1929.

[3] I. e. A not O, B not O.

The conclusions below follow at once:

(a) $d(A, B) = 0$ if and only if $A = B$.

(b) if any three distinct points A, B, C of a line are properly ordered the equation

$$d(A, C) = d(A, B) + d(B, C)$$

holds, but in this order A, B, C and the inverse order C, B, A only.

(c) if x' denotes any second system of numeration in the line l then for all points A of l and for some constant d either $x'_A = x_A + d$ or else $x'_A = -x_A + d$.

Clearly (c) affirms that the numeration on the scale is determined by the origin marked 0 and the *direction* taken as positive.

We may now define a point B as *between* A and C ($A \neq C$) if the relation above written holds. The points A and C together with all points B between A and C forms the *segment* AC: in other words the segment AC is the set of points P such that $\begin{Bmatrix} x_A \leqq x_P \leqq x_C \\ x_C \leqq x_P \leqq x_A \end{Bmatrix}$ according as $\begin{Bmatrix} x_A < x_C \\ x_C < x_A \end{Bmatrix}$.

Likewise a *half-line* l' with *end-point* O is defined by two points O, A in a line $l (A \neq O)$ as the class of all points A' of l such that O is not between A and A'; in other words if O is taken as the origin of the system of numeration and A is taken on the positive side, i. e. $x_A > 0$, the half-line OA consists precisely of the points P for which $x_P \geqq 0$. The given line l consists of two such half-lines with given common end-point O.

Fig. 4.

If A, B, C are three distinct points the three segments AB, BC, CA are said to form a *triangle* $\triangle ABC$, with *sides* AB, BC, CA and *vertices* A, B, C. If A, B, C are in the same straight line, $\triangle ABC$ is said to be *degenerate*, otherwise *non-degenerate*.

3. The point-line postulate. This second postulate is stated as follows:

POSTULATE II. *One and only one straight line l contains two given points P, Q ($P \neq Q$).*

In consequence of II, any two distinct lines l, m have either one point in common or none. In the first case they are said to *intersect* in their common point; in the second case, they are said to be *parallel*;[4] a line l is always regarded as parallel to itself.

[4] True only for the geometry of the plane of course.

Postulate II embodies the fact that a unique line may be passed through two points, so that two lines intersect, if at all, in a determinate point. It is upon this fact that all geometric constructions depend.

4. **The postulate of angle measure.** The third postulate of angle measure may be stated as follows:

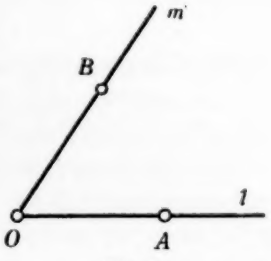

Fig. 5.

POSTULATE III. *The half-lines l, m, $\cdots$, through any point O can be put into* (1, 1) *correspondence with the real numbers a (mod 2π), so that, if $A \neq O$ and $B \neq O$ are points of l and m respectively, the difference $a_m - a_l \pmod{2\pi}$ is $\angle AOB$.*[5] *Furthermore, if the point B on m varies continuously in a line r not containing the vertex O, the number a_m varies continuously also.*[6]

According to this postulate it is apparent that any two half-lines l, m with common end-point O define an angle $\angle lOm$, namely $\angle AOB$ where $A \neq O$ and $B \neq O$ lie in l and m respectively.

Evidently the protractor embodies this postulate.

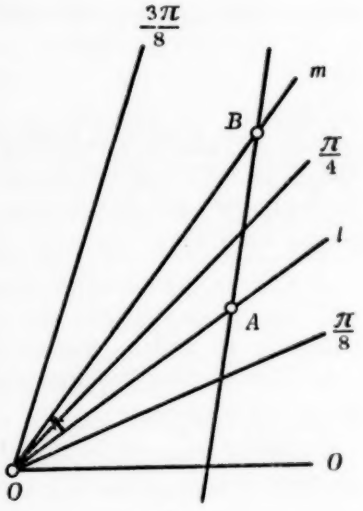

Fig. 6.

It will be seen that the angle $\angle lOm$ *as here conceived* is the directed angle from the half-line l to the half-line m *determining the position of m relative to l*. The *ordinary* angle $\angle lOm$ is then given by the numerical value of the least residue of $a_m - a_l \pmod{2\pi}$. The *ordinary sensed* angle of the usual type is obtained by taking some single algebraic difference $a_m - a_l$ which is thought of as representative of an angle generated by the continuous rotation of a half-line from l to m.

---

[5] According to the usual mathematical notation the symbol $d \pmod{2\pi}$, read d, modulo 2π, stands for the infinite set $d + 2k\pi$, $k = 0, \pm 1, \pm 2, \cdots$, having the same residue as d when divided by 2π. The least of these numbers does not exceed π in absolute value and is called the least residue.

[6] More precisely, $\lim a_m = a_l$ if $\lim_{B=A} d(B, A) = 0$ for points B, A of such a line m (see Fig. 6). It is the second part of Postulate III that excludes "non-Archimedean" possibilities: from the elementary point of view this second part may well be omitted in a first approach to geometry.

The following conclusions appear at once:

(a) $\angle lOm \equiv 0$ if and only if $l = m$.

(b) if l, m, n are any three half-lines through O, then

$$\angle lOm + \angle mOn \equiv \angle lOn.$$

(c) if a' denotes any second system of numeration, then for all half-lines l through O and for some constant d either $a_l' \equiv a_l + d$, or else $a_l' \equiv -a_l + d$.

Two half-lines l, m through O are said to form a *straight angle* if

$$\angle lOm \equiv \pi.$$

It is clear that if $\angle lOm$ is a straight angle so is $\angle mOl$, according to this definition.

Two half-lines l, m through O are said to form a *right* angle if

$$\angle lOm \equiv \pm\frac{\pi}{2}.$$

It is clear that if $\angle lOm$ is a right angle so is $\angle mOl$. We say also that m is perpendicular to l, written $m \perp l$, in this case. It follows that there are two half-lines $m_1, m_2 \perp$ to a given half-line l at its end-point O, namely such that

$$\angle lOm_1 \equiv \frac{\pi}{2}, \qquad \angle lOm_2 \equiv -\frac{\pi}{2}.$$

Evidently m_1 and m_2 form a straight angle.

5. **The similarity postulate.** The fourth and last postulate of similarity is the following:

Postulate IV. *If in two triangles,* $\triangle ABC$, $\triangle A'B'C'$, *and for some constant* $k>0$, $d(A', B') = kd(A, B)$, $d(A', C') = kd(A, C)$ *and also* $\angle B'A'C' = \pm\angle BAC$ *then also* $d(B', C') = kd(B, C)$, $\angle C'B'A' = \pm\angle CBA$, $\angle A'C'B' = \pm\angle ACB$.

Fig. 7a.

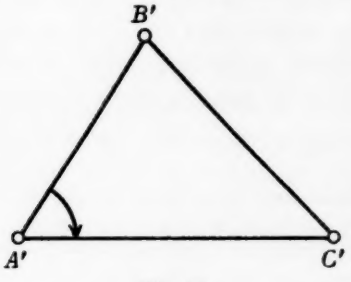

Fig. 7b.

In other words if we define two triangles as *similar* in case corresponding sides are proportional, and corresponding angles are either all equal, or all the negatives of one another, this postulate states that if $\triangle ABC$ and $\triangle A'B'C'$ have two sides proportional and the corresponding included angle equal, they are similar.

A *broken line* $ABC\cdots KL$ consists of a collection of segments AB, $BC, \cdots, KL$, which may or may not intersect in certain points. The points $A, B, \cdots, L$ are the *vertices* of the broken line. If the initial point A and the terminal point L coincide, the broken line is called a *polygon.*

Two broken lines or polygons are said to be similar if corresponding sides are in proportion and corresponding angles are all equal or all the negatives of one another. It follows immediately from Postulate IV that if only $A'B', \cdots, K'L'$ and $AB, \cdots, KL$ respectively are in proportion while $\angle A'B'C', \cdots, \angle J'K'L'$ and $\angle ABC, \cdots, \angle JKL$ respectively are equal or the negatives of one another, such broken lines or polygons must be similar.

More generally any two geometric figures whatsoever may be termed *similar* if there exists a correspondence of points such that all corresponding distances are in proportion and corresponding angles are all equal or all the negatives of one another.

In an analogous manner two broken lines or triangles or polygons or figures may be said to be *congruent* if they are similar with ratio of proportionality $k = 1$ of distances, so that corresponding distances are equal.

It will be observed that Postulate IV applies in the degenerate case when the points A, B, C or A', B', C' lie in a straight line, since $ABC, A'B'C'$ may form (degenerate) triangles in that case also, according to definition.

Evidently this last postulate is essentially an embodiment of the fact that any geometric figure may be diminished or enlarged in an arbitrary ratio k with prescribed arbitrary orientation in the plane.

In showing how these postulates suffice as a basis for plane geometry, we shall prove certain of the simplest and most fundamental theorems, I–X, from which it follows at once that this set of postulates characterizes the Euclidean plane categorically. With this material in hand further geometric developments follow as usual of course.

6. **The theorem on straight angles.** A first consequence of Postulates I–IV is given by the following theorem:

THEOREM I. *The angle lOm is a straight angle if and only if l and m are the two half-lines of a single line n, which have O as common end-point.*[7]

Proof. Suppose first that n is a line with point O and corresponding half-lines l and m. Let us show that $\angle lOm \equiv \pi$.

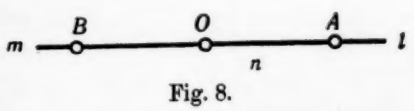

Fig. 8.

[7] From the elementary point of view Theorem I may well be presented as an independent postulate complementary to Postulate III.

Choose A in l and B in m so that $OA = OB = 1$ (Postulate I). In degenerate $\triangle OBA$ and $\triangle OAB$ we have

$$\angle OBA \equiv +\angle OAB \equiv 0$$

and also

$$d(B, O) = d(A, O), \qquad d(B, A) = d(A, B)$$

so that $\triangle OBA$ and $\triangle OAB$ are congruent, i. e. similar with $k = 1$, by Postulate IV and the remaining corresponding angles are the same:

$$\angle BOA \equiv \angle AOB \equiv -\angle BOA$$

whence

$$2\angle BOA \equiv 0.$$

It follows that either $\angle BOA \equiv 0$ or π. But the possibility $\angle BOA \equiv 0$ must be excluded since OA and OB are distinct half-lines (Postulate III). Hence we infer $\angle BOA \equiv \pi$ as desired.

Suppose secondly that the half-lines l and m meet at O to form a straight angle π. The other half-line l' in the same straight line as l with end-point O also forms a straight angle π with l by what precedes, so that

$$\angle lOm \equiv \pi, \qquad \angle l'Ol \equiv \pi.$$

Hence

$$\angle l'Om \equiv \angle l'Ol + \angle lOm \equiv 0,$$

i. e. m must coincide with l'.

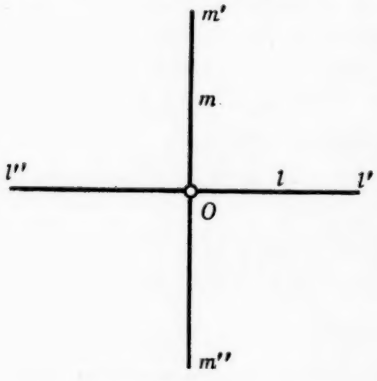

Fig. 9.

From the above theorem it appears at once that two unlimited straight lines l and m are appropriately called perpendicular, $l \perp m$, if they meet at a point O with perpendicular half-lines l', l'' and m', m''.

7. The three triangle theorems. Let us next prove the following familiar theorem:

THEOREM II. *Two triangles $\triangle ABC$ and $\triangle A'B'C'$ are similar if two pairs of corresponding angles are equals or negatives of one another.*

Proof. Suppose for example $\angle ABC \equiv \angle A'B'C'$ and $\angle CAB \equiv \angle C'A'B'$. Define $k = \dfrac{A'B'}{AB}$ and let C'' be the point on the half-line $A'C'$ such

that $\frac{A'C''}{AC} = k$ also. Clearly by Postulate IV, $\triangle A'B'C''$ is similar to $\triangle ABC$ with equal angles. Hence $\angle A'B'C'' \equiv \angle ABC \equiv \angle A'B'C'$. By Postulate III, $B'C'$ must then be the same half-line as $B'C''$. Hence

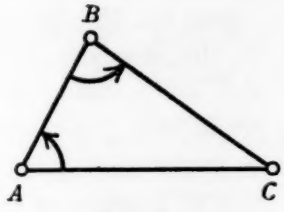

Fig. 10a.

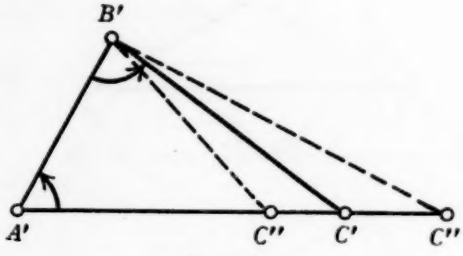

Fig. 10b.

C'' lies in the line $B'C'$ as well as in $A'C'$ and so must coincide with C' by Postulate II. Obviously then $\triangle A'B'C'$ is necessarily similar to $\triangle ABC$, as stated.

Secondly let us prove:

THEOREM III. *If* $d(A, C) = d(B, C)$ *in the* $\triangle ABC$, *then* $\angle CAB \equiv -\angle CBA$; *and conversely.*

Proof. Compare triangles $\triangle CAB$ and $\triangle CBA$ in which $d(C, A) = d(C, B)$, $d(C, B) = d(C, A)$, $\angle ACB \equiv -\angle BCA$. These triangles are congruent (Postulate IV with $k = 1$), and $\angle CAB \equiv -\angle CBA$.

To prove the converse, we assume $\angle CAB \equiv -\angle CBA$ and compare the triangles $\triangle CAB$ and $\triangle CBA$ again in which also

$$\angle BCA \equiv -\angle ACB.$$

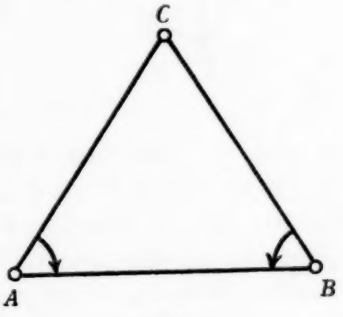

Fig. 11.

Hence by Th. II above, the triangles are similar with $k = 1$. It follows that $d(A, C) = d(B, C)$ as stated.

A triangle with two equal sides is called *isosceles.*

Thirdly let us prove:

THEOREM IV. *Two triangles* $\triangle ABC$ *and* $\triangle A'B'C'$ *are similar if their corresponding sides are proportional.*

Proof. Construct the $\triangle A'B'C''$ similar to $\triangle ABC$ and so that $\angle C''A'B'$ is of opposite sign[8] to $\angle C'A'B'$ (Postulate IV). This is always possible

[8] That is, the least residues, mod 2π, are of opposite sign.

unless $\angle C'A'B'$ or $\angle CAB \equiv 0$ or π. For the moment we exclude such cases so that C' and C'' are distinct. It will obviously suffice to prove that the $\triangle A'B'C'$ and $\triangle A'B'C''$ are similar; it is clear that their corresponding sides are equal.

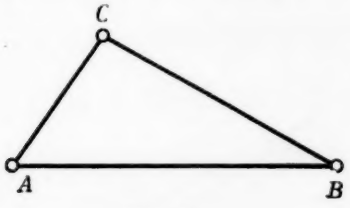

Fig. 12a.

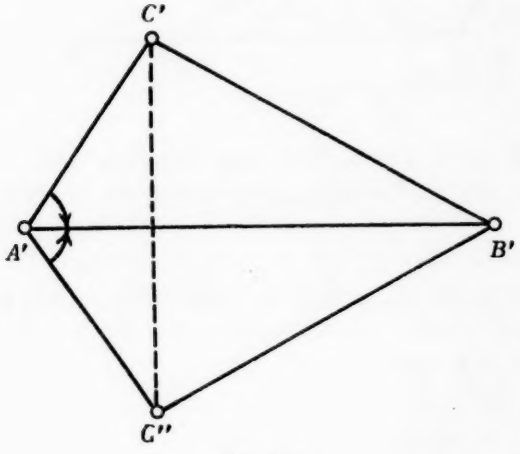

Fig. 12b.

But by Th. III, applied to the isosceles triangles, $\triangle C'A'C''$ and $\triangle C'B'C''$, we have

$$\angle A'C''C' \equiv -\angle A'C'C'',$$
$$\angle C'C''B' \equiv -\angle C''C'B',$$

whence by addition

$$\angle A'C'B' \equiv -\angle A'C''B'.$$

Thus the angles in $\triangle A'B'C'$ and $\triangle A'B'C''$ with vertices at C' and C'' are the negatives of one another, and these triangles are similar.

Evidently the same proof applies also in the exceptional cases so long as we can take $C'' \neq C'$. But such a position of C'' will be obtained unless we have $\angle CAB \equiv \angle C'A'B' \equiv 0$ or π. In this case the theorem stated is obviously true by Postulate IV. Thus the proof is completed.

Postulate IV and Ths. II, IV form the usual bases for congruence and similarity theorems, which we need not develop further.

8. **The angle-sum theorem.** The following theorem can now be established at once:

THEOREM V. *In any triangle* $\triangle ABC$: $\angle ABC + \angle BCA + \angle CAB \equiv \pi$. *Furthermore if* $\triangle ABC$ *is non-degenerate all three angles may be taken of the same sign between* 0 *and* π.[9]

Proof. To prove the first statement let the mid-points of the sides BC, CA, AB be K, L, M respectively. We observe first that, according to Postulate IV, $\triangle AML$, $\triangle MBK$, $\triangle LKC$ are similar to $\triangle ABC$ with $k = \frac{1}{2}$ and equal angles. It follows that $d(L, M) = \frac{1}{2} d(B, C)$,

[9] The second part of this theorem may be regarded as self-evident from the elementary point of view.

$d(M, K) = \frac{1}{2}d(C, A)$, $d(K, L) = \frac{1}{2}d(A, B)$. Thus the $\triangle KLM$ is congruent to each of the small triangles by Th. IV. If then we write

$$\angle ABC \equiv \beta, \quad \angle BCA \equiv \gamma, \quad \angle CAB \equiv \alpha,$$

we have

$$\angle MKL \equiv \pm\alpha, \quad \angle KLM \equiv \pm\beta, \quad \angle LMK \equiv \pm\gamma$$

where the $+$ or $-$ sign is to be the same throughout.

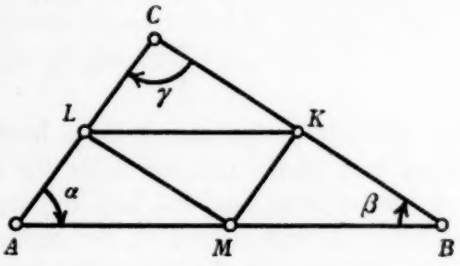

Fig. 13.

Now the angle

$$\angle AMB \equiv \angle AML + \angle LMK + \angle KMB \equiv \beta + \angle LMK + \alpha$$

is a straight angle. We obtain then the desired relation if the $+$ sign holds.

However, in case the $-$ sign holds, we have

$$\pi \equiv \beta - \gamma + \alpha \qquad \text{and similarly} \qquad \pi \equiv \gamma - \alpha + \beta,$$

$$\pi \equiv \alpha - \beta + \gamma,$$

whence, by addition

$$3\pi \equiv \pi \equiv \alpha + \beta + \gamma$$

(as before) so that from the above equations

$$2\alpha \equiv 2\beta \equiv 2\gamma \equiv 0,$$

and therefore

$$\alpha \equiv 0 \text{ or } \pi, \quad \beta \equiv 0 \text{ or } \pi, \quad \gamma \equiv 0 \text{ or } \pi.$$

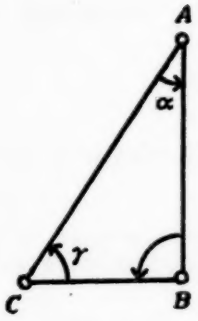

Fig. 14.

Thus the $+$ sign holds here and so in all cases.

To establish the second part we proceed as follows:

The second part holds for any *right-angle* $\triangle ABC$ with right angle at B. In fact we may order A, B, C so that $\angle ABC \equiv \frac{\pi}{2}$. Let now C tend toward B along the line BC. The $\angle CAB$ varies continuously (Postulate III, part 2) and never takes on the value 0, π or $\frac{\pi}{2} \pmod{2\pi}$, since C never lies in the line AB and since $\angle BCA \not\equiv 0$ (use Th. V, part 1). Hence the least residue α of $\angle CAB$ must be less

numerically than $\frac{\pi}{2}$. Similarly the least residue γ of $\angle BCA$ is less numerically than $\frac{\pi}{2}$. Now we have

$$\alpha+\frac{\pi}{2}+\gamma \equiv \pi \quad \text{or} \quad \alpha+\gamma \equiv \frac{\pi}{2}.$$

This can only be the case if these least residues are positive. Hence the second part of the theorem holds for any right triangle.

A similar reasoning will hold for any $\triangle ABC$ not a right triangle unless in letting B tend toward C, for instance, $\angle CAB$ becomes a right angle at D. In fact, if this does not happen for some variation we should conclude that the least residues α, β, γ of $\angle CAB, \angle ABC, \angle BCA$ were numerically less than $\frac{\pi}{2}$. But then we should have also

$$\alpha+\beta+\gamma \equiv \pi$$

so that the least residues are necessarily all of the same sign (+ or —), and the conclusion is the same as before.

But if there is such a point D we see at once by a consideration of the right triangle $\triangle ADB$ that $\angle BAD$ and $\angle DBA$ can be taken as $<\frac{\pi}{2}$ numerically and of the same sign as $\angle ADB \equiv \pm\frac{\pi}{2}$, with

$$\angle BAD+\angle DBA \equiv \pm\frac{\pi}{2}.$$

Likewise from the right triangle $\triangle CDA$ we obtain

$$\angle ACD+\angle DAC \equiv \angle CDA \equiv \pm\frac{\pi}{2}.$$

However since we have

$$\angle ADB+\angle CDA \equiv \angle CDB \equiv \pi$$

the same $\pm$ sign appears in both cases. By addition then

$$\angle BAC+\angle ACB+\angle CBA \equiv \pm\pi$$

where the least residues in the left member are of the same $\pm$ sign and less than π numerically.

Of course only one residue can exceed $\frac{\pi}{2}$.

9. **A property of the perpendicular bisector.** The *perpendicular bisector* of a line segment AB is the line $\perp AB$ at the mid-point D of AB,

i. e. the point D for which $d(A, D) = d(D, B)$. By Postulate I, such a point exists. The perpendicular bisector has the following important "locus" property:

THEOREM VI. *All points P equidistant from A and B ($A \neq B$) and no others lie on the perpendicular bisector of the line AB.*

Proof. Given A, B and a point P such that $d(P, A) = d(P, B)$. Let M be the mid-point of AB and join P to M. The two triangles $\triangle AMP$ and $\triangle BMP$ are then congruent by Th. IV, since

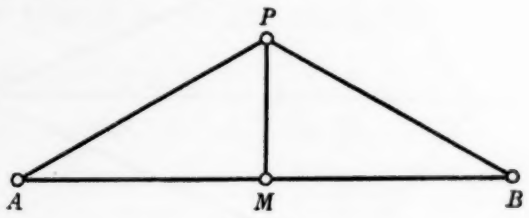

Fig. 15.

$$d(A, P) = d(B, P), \qquad d(A, M) = d(B, M), \qquad d(M, P) = d(M, P).$$

Hence $\angle MPB \equiv \pm \angle MPA$. But the $-$ sign must be taken since A and B are distinct. Hence also

$$\angle BMP \equiv -\angle AMP \equiv \angle PMA.$$

Therefore

$$\angle AMB \equiv \angle AMP + \angle PMB \equiv 2\angle AMP$$

or

$$\pi \equiv 2\angle AMP.$$

In consequence

$$\angle AMP \equiv \pm\frac{\pi}{2},$$

i. e. such a point P lies in the perpendicular bisector of AB.

Moreover, suppose P is *any* point in the perpendicular bisector, then in the triangles $\triangle AMP$, $\triangle BMP$

$$\angle AMP \equiv -\angle BMP, \quad d(A, M) = d(M, B), \quad d(M, P) = d(M, P).$$

Hence by Postulate IV the triangles are congruent, and accordingly $d(A, P) = d(B, P)$. This completes the proof.

10. **The existence of a unique perpendicular from a point to a line.** We prove next the following theorem:

THEOREM VII. *There is one and only one line perpendicular to a line l which contains a given point P.*

Proof. For P on l the statement is obviously true.

For P not on l we proceed as follows. Let A, B be two distinct points of l. Construct a $\triangle AP'B$, congruent to $\triangle APB$ with P' distinct

from P; this is done by choosing a ray AP' so that $\angle P'AB \equiv -\angle PAB$ and taking P' so that $d(A, P) = d(A, P')$, thus determining $\triangle AP'B$. Here $P \neq P'$ of course. But A and B are equally distant from P and P' and so lie on the perpendicular bisector of PP'. Thus the line PP' and the line AB are $\perp$, and a perpendicular to AB through P is thus actually obtained.

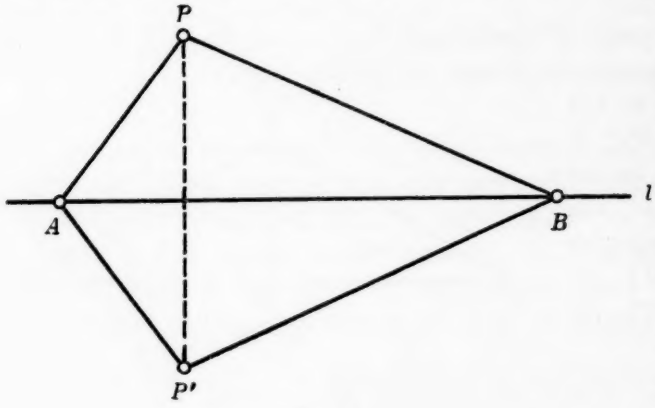

Fig. 16.

If there were a second perpendicular to AB through P, we would have a triangle PMM' where PM and PM' were the perpendiculars in question with M and M' on l. But, by Th. V,

$$\pi \equiv \angle MM'P + \angle M'PM + \angle PMM'$$

or

$$\pi \equiv \pm\frac{\pi}{2} + \angle M'PM \pm \frac{\pi}{2}$$

whence

$$\angle M'PM \equiv 0 \text{ or } \pi,$$

i. e. $M'P$ and MP would be the same straight line, contrary to hypothesis.

11. The Pythagorean theorem. We may now prove the Pythagorean theorem:

THEOREM VIII. *In any right* $\triangle ABC$ *with* $\angle ACB \equiv \pm\frac{\pi}{2}$,

$$d(A, B)^2 = d(A, C)^2 + d(C, B)^2$$

or $c^2 = a^2 + b^2$.

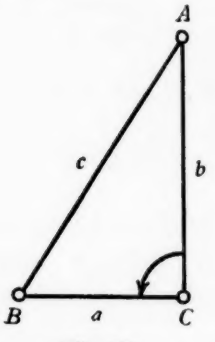

Fig. 17.

Proof. If we enlarge this triangle in the ratio b to 1 (Postulate IV) it becomes a right triangle with sides ab, b^2, bc similar to the first with $k = b$ and the same angle $\angle A'C'B'$ in the new triangle $A'B'C'$ so obtained. Similarly a triangle $\angle A''B''C''$ with $k = a$ has sides a^2, ab, ac respectively

and $\angle A''C''B''$ as before. If these triangles be "fitted together" (see Fig. 18) so that B', C' and A'', C'' respectively are the ends of a given line segment, it is clear that

$$\angle A'C'B'' \equiv \angle A'C'B' + \angle B'C'B'' \equiv \pm\frac{\pi}{2} \pm\frac{\pi}{2} \equiv \pm\pi$$

so that C' is on the line $A'B''$ between A' and B''. Furthermore

$$\angle B''B'A' \equiv \angle B''B'C' + \angle C'B'A'$$
$$\equiv \angle BAC + \angle CBA \equiv \pm\frac{\pi}{2} \text{ by Th. V.}$$

It follows that $\triangle A'B''B'$ is similar to $\triangle ABC$ with angles reversed in sign.

But on the other hand it is clear that the scale factor k involved is $k = c$ since the distances $d(A, C)$, $d(C, B)$ become bc and ac in the new triangle. The third side is therefore c^2. By comparison then

$$c^2 = a^2 + b^2 \text{ (Postulate I).}$$

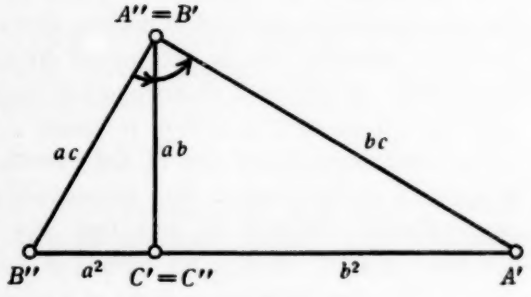

Fig. 18.

12. **Parallels. Rectangular networks.** A fundamental property of parallel lines in the plane is the following:

THEOREM IX. *One and only one line parallel to a given line contains a given point P.*

Proof. If P is on l, this fact is obvious, l being parallel to itself.

If P is not on l, consider the unique line $PD \perp l$, and the line $m \perp PD$ at P.

This line m cannot intersect l, since otherwise there would be two perpendiculars from the point of intersection to PD, meeting it in P and D, in contradiction with Th. VII.

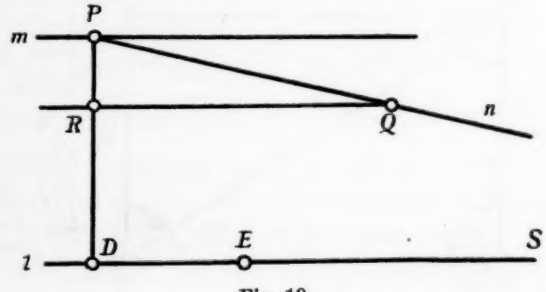

Fig. 19.

Hence the line m is $\parallel$ to l. It remains to show that it is the only such line through P.

If there were another line, n, choose a point $Q \neq P$ on n, and from Q construct the perpendicular to PD meeting it in R. Lay off

on the half-line DE of l such that $\angle QRP \equiv \angle EDP \equiv \pm\frac{\pi}{2}$ a distance DS such that

$$\frac{SD}{PD} = \frac{QR}{PR}.$$

The $\triangle PDS$ is then similar to $\triangle PRQ$ with *equal* angles (Postulate IV). Hence the half-line PS coincides with PQ and the line n is not $\parallel l$, since it meets l in S. This is contrary to assumption.

It follows at once from Th. IX that if l, m, n are any three lines such that $l \parallel m$ and $m \parallel n$, then $l \parallel n$ also. For if l intersects n there would be two parallels to m through this point of intersection.

The set of the lines parallel to a given line l and so to each other is called a *system of parallels*, and is determined by any one of its lines. One and only one line of any such system passes through any point of the plane.

Any line not in such a system is called a *transversal* of the system and of course intersects every one of the system of parallel lines.

It appears at once that a line perpendicular to one line of such a system is perpendicular to them all and that any two such perpendiculars are themselves parallel. Hence we obtain a *rectangular network* formed by two mutually perpendicular systems of parallels in which one and only one line of each system passes through any point. Clearly such a network is uniquely defined by any single line l of either system.

Two pairs of lines of each system l_1, l_2 and m_1, m_2 intersect in four distinct points:

$$l_1, m_1 \text{ in } P; \quad l_1, m_2 \text{ in } S; \quad l_2, m_1 \text{ in } Q; \quad l_2, m_2 \text{ in } R.$$

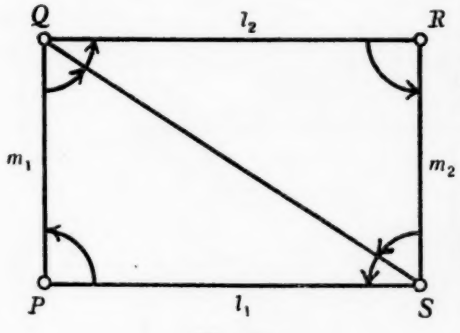

Fig. 20.

The polygon of four sides $PQRS$ is a *rectangle* by definition since the four angles at P, Q, R, S are right angles.

THEOREM X. *In any rectangle $PQRS$ the two pairs of opposite sides PQ, RS and PS, RQ are equal, and the four angles $\angle PQR$, $\angle QRS$, $\angle RSP$, $\angle SPQ$ are all congruent to $\frac{\pi}{2}$ or to $-\frac{\pi}{2}$.*

Proof. The truth of the first part of the theorem is an immediate consequence of the application of Postulate IV to a pair of triangles such as $\triangle SPQ$, $\triangle QRP$, since these are evidently congruent right triangles, so that $PS = QR$.

The second part of the theorem may be proved as follows:

By the angle-sum Th. V applied to the triangles $\triangle SPQ$, $\triangle QRP$ it appears that the sum of the four right angles is $\equiv 2\pi$. Moreover by the same theorem $\angle RSQ$ and $\angle SQR$ are numerically $< \frac{\pi}{2}$ and of the same sign as the right angle $\angle QRS$; likewise $\angle QSP$ and $\angle PQS$ are numerically $< \frac{\pi}{2}$ and of the same sign as the right angle $\angle SPQ$. But we have

$$\angle RSQ + \angle QSP \equiv \pm \frac{\pi}{2}$$

so that the least residues of $\angle RSQ$ and $\angle QSP$ must have the same sign. Consequently the least residues of all the angles written have the same sign $\pm$, and the four right angles of the rectangle are all congruent to $+\frac{\pi}{2}$ or to $-\frac{\pi}{2}$.

From this theorem it follows that we may speak of a sense or positive direction on the lines of a system of parallels, namely that obtained by attaching the same numbers x_R to a point R of l as to the corresponding points S of intersection with any other line m of the system, i. e. $x_S = x_R$.

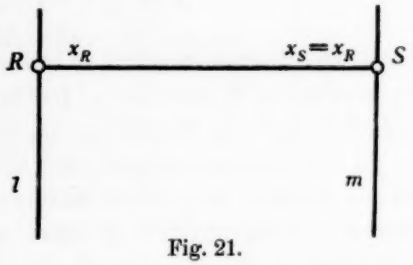

Fig. 21.

In general when we speak of a system of parallels we refer to a *directed system* of this kind. Similarly we think of a rectangular network as *directed*, i. e. of each of its component systems as directed.

It follows then from the above theorem that *in a directed rectangular network the directed lines l of one system make with the directed lines m of the other system precisely the same angle* $\frac{\pi}{2}$ *or* $-\frac{\pi}{2}$ *in all cases.*

At this stage the usual theorems concerning parallel lines may now be proved at once, for instance the fact that *a directed transversal n of a system of directed parallels intersects all of them at the same angle.* In general the ordinary theorems of Euclidean geometry not yet deduced follow as easy exercises from the results so far obtained and it is unnecessary to proceed further in this direction.

13. **Rectangular coördinates.** We are now prepared to define *a rectangular coördinate system (x, y) with axes* the directed perpendicular lines Ox, Oy intersecting in the origin O.

Choose a system of numeration along Ox, Oy so that O is marked 0 on both lines, and the numbers increase algebraically in the positive direction. Drop the unique perpendiculars from P to Ox and Oy respectively meeting these lines at P_x and P_y. *The rectangular coördinates of P are then the numbers x_P and y_P attached to P_x and P_y respectively.* These are evidently uniquely determined numbers.

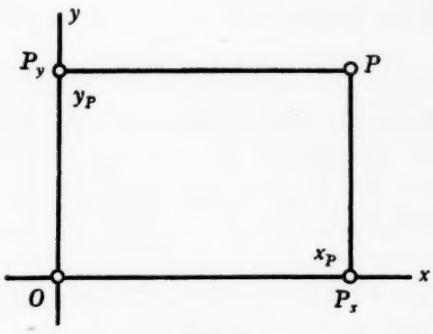

Fig. 22.

It may now be readily proved on the basis of the theorems established above that the "equation of any straight line" has the form

$$Ax + By + C = 0$$

where not both A and B are 0, and that conversely any equation of this form represents a straight line.

Furthermore the Pythagorean theorem shows that

$$d(P, Q) = \sqrt{(x_P - x_Q)^2 + (y_P - y_Q)^2}.$$

14. **Euclidean angle.** In order to show that the above set of postulates is categorical it remains to indicate briefly why our undefined angle must coincide with the Euclidean angle (mod 2π).

To this end we observe that the circle is defined as usual as the locus of all points P at a fixed distance r (the radius) from a fixed point O (the center). Evidently if O is taken as the origin of a system of rectangular coördinates, the equation of the circle is

$$x^2 + y^2 = r^2.$$

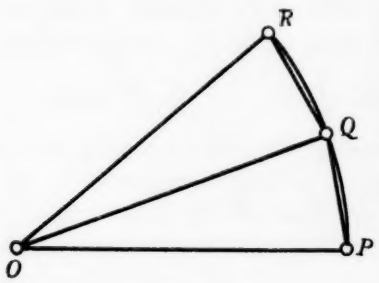

Fig. 23.

On the basis of the preceding theorems, Euclidean arc length can be defined in the usual manner and the angle POQ (P, Q on the circle) may be defined as the arc PQ on the unit circle ($r = 1$). Furthermore it is not hard to prove on the basis of Postulate III, part 2, that $\angle POQ$ varies continuously with continuous variation of either P or Q.

Consider now the ratio $\dfrac{\angle POQ}{\text{arc } PQ}$ for a fixed position of P and for variable Q. It is easy to prove that if arc PQ is multiplied in any

integral ratio $m:1$ so is $\angle POQ$ multiplied in the same ratio $m:1$. For example, if the arc length is doubled so that arc PQ = arc QR, then $d(P, Q) = d(Q, R)$, of course, so that $\angle POQ \equiv \angle QOR$ and $\angle POR \equiv \angle POQ + \angle QOR \equiv 2\angle POQ$. Hence the stated results hold for $m = 2$

$$\frac{\angle POR}{\text{arc } PR} = \frac{\angle POQ}{\text{arc } PQ}.$$

The same type of proof can be extended obviously for $m = 3, 4, \cdots$.

Inversely it follows that if arc PQ be divided in the ratio $1:n$ so is $\angle POQ$ divided in this ratio. Consequently if the ratio has a value k for arc $PQ = s_0$, it has the same value for arc $PQ = \frac{m}{n} s_0$, where m and n are any positive integers whatever. In virtue of the second part of Postulate III, it follows, however, that the ratio is a continuous function of s so that $\frac{\angle POQ}{\text{arc } PQ}$ is a constant k for all $Q \neq P$. But for arc $PQ = \pi$, we have $\angle POQ = \pi$ also so that this ratio is 1. Hence the (sensed) angle $\angle POQ$ coincides with the (sensed) arc length PQ subtended on the unit circle.

It follows that Postulates I–IV form a categorical system.

A SET OF POSTULATES FOR THE FOUNDATION OF LOGIC.[1]

By Alonzo Church.[2]

1. **Introduction.** In this paper we present a set of postulates for the foundation of formal logic, in which we avoid use of the free, or real, variable, and in which we introduce a certain restriction on the law of excluded middle as a means of avoiding the paradoxes connected with the mathematics of the transfinite.

Our reason for avoiding use of the free variable is that we require that every combination of symbols belonging to our system, if it represents a proposition at all, shall represent a particular proposition, unambiguously, and without the addition of verbal explanations. That the use of the free variable involves violation of this requirement, we believe is readily seen. For example, the identity

$$a(b+c) = ab+ac \tag{1}$$

in which a, b, and c are used as free variables, does not state a definite proposition unless it is known what values may be taken on by these variables, and this information, if not implied in the context, must be given by a verbal addition. The range allowed to the variables a, b, and c might consist of all real numbers, or of all complex numbers, or of some other set, or the ranges allowed to the variables might differ, and for each possibility equation (1) has a different meaning. Clearly, when this equation is written alone, the proposition intended has not been completely translated into symbolic language, and, in order to make the translation complete, the necessary verbal addition must be expressed by means of the symbols of formal logic and included, with the equation, in the formula used to represent the proposition. When this is done we obtain, say,

$$R(a)\,R(b)\,R(c)\,\supset_{abc} \cdot\, a(b+c) = ab+ac \tag{2}$$

where $R(x)$ has the meaning, "x is a real number," and the symbol $\supset_{abc}$ has the meaning described in §§ 5 and 6 below. And in this expression there are no free variables.

---

[1] Received October 5, 1931.

[2] This paper contains, in revised form, the work of the author while a National Research Fellow in 1928-29.

A further objection to the use of the free variable is contained in the duplication of symbolism which arises when the free, or real, variable and the bound, or apparent, variable are used side by side.[3] Corresponding to the proposition, represented by equation (1) when a, b, and c stand for any three real numbers, there is also a proposition expressed without the use of free variables, namely (2), and between these two propositions we know of no convincing distinction. An attempt to identify the two propositions is, however, unsatisfactory, because substitution of (1) for (2), when the latter occurs as a part of a more complicated expression, cannot always be allowed without producing confusion. In fact, the only feasible solution seems to be the complete abandonment of the free variable as a part of the symbolism of formal logic.[4]

Rather than adopt the method of Russell for avoiding the familiar paradoxes of mathematical logic,[5] or that of Zermelo,[6] both of which appear somewhat artificial, we introduce for this purpose, as we have said, a certain restriction on the law of excluded middle. This restriction consists in leaving open the possibility that a propositional function F may, for some values X of the independent variable, represent neither a true proposition nor a false proposition. For such a value X of the independent variable we suppose that $\mathsf{F}(\mathsf{X})$ is undefined and represents nothing, and we use a system of logical symbols capable of dealing with propositional functions whose ranges of definition are limited.

In the case of the Russell paradox the relevance of this proposed restriction on the law of excluded middle is evident. The formula P which leads to this paradox may be written, in the notation explained below, $\{\lambda\varphi . \sim\varphi(\varphi)\}\,(\lambda\varphi . \sim\varphi(\varphi))$. It has the property that if we assume $\sim\mathsf{P}$ then we can infer P and if we assume P then we can infer $\sim\mathsf{P}$. On ordinary assumptions both the truth and the falsehood of P can be deduced in consequence of this property, but the system of this paper, while it provides for the existence of a propositional function $\lambda\varphi . \sim\varphi(\varphi)$ does not provide either that this propositional function shall be true or that it shall be false, for the value $\lambda\varphi . \sim\varphi(\varphi)$ of the independent variable.

Other paradoxes either disappear in the same way, or else, as in the case of the Epimenides or the paradox of the least undefinable ordinal,

[3] Cf. the introduction to the second edition of Whitehead and Russell's *Principia Mathematica*.

[4] Unless it is retained as a mere abbreviation of notation.

[5] B. Russell, Mathematical Logic as based on the Theory of Types, Amer. Jour. Math., vol. 30 (1908), pp. 222–262. A list of some of these paradoxes, with a reference to the source of each, will be found in this article, or in Whitehead and Russell, *Principia Mathematica*, vol. 1, pp. 63–64.

[6] E. Zermelo, Untersuchungen über die Grundlagen der Mengenlehre, Math. Annalen, vol. 65 (1908), pp. 261–281.

they contain words which are not definable in terms of the undefined symbols of our system, and hence need not concern us.

The paradox of Burali-Forti is not, however, so readily disposed of. The question whether this paradox is a consequence of our postulates, or what modification of them will enable us to avoid it, probably must be left open until the theory of ordinal numbers which results from the postulates has been developed.

Whether the system of logic which results from our postulates is adequate for the development of mathematics, and whether it is wholly free from contradiction, are questions which we cannot now answer except by conjecture. Our proposal is to seek at least an empirical answer to these questions by carrying out in some detail a derivation of the consequences of our postulates, and it is hoped either that the system will turn out to satisfy the conditions of adequacy and freedom from contradiction or that it can be made to do so by modifications or additions.

2. **Relation to intuitionism.** Since, in the postulate set which is given below, the law of the excluded middle is replaced by weaker assumptions, the question arises what the relation is between the system of logic which results from this set, and the intuitionism of L. E. J. Brouwer.[7]

The two systems are not the same, because, although both reject a certain part of the principle of the excluded middle, the parts rejected are different. The law of double negation, denied by Brouwer, is preserved in the system of this paper, and the principle, which Brouwer accepts, that a statement[8] from which a contradiction can be inferred is false, we find it necessary to abandon in certain cases.

Our system appears, however, to have the property, which relates it to intuitionism, that a statement of the form $\Sigma x . F(x)$ (read, "there exists x such that $F(x)$") is never provable unless there exists a formula M such that $F(M)$ is provable.

3. **The abstract character of formal logic.** We do not attach any character of uniqueness or absolute truth to any particular system of logic. The entities of formal logic are abstractions, invented because of their use in describing and systematizing facts of experience or observation, and their properties, determined in rough outline by this intended use, depend for their exact character on the arbitrary choice of the inventor.

---

[7] See L. E. J. Brouwer, Intuitionistische Mengenlehre, Jahresbericht der D. Math. Ver., vol. 28 (1919), pp. 203–208, and Mathematik, Wissenschaft und Sprache, Monatshefte für Math. u. Phys., vol. 36 (1929), pp. 153–164, and many other papers.

[8] We purposely use the word, "statement", because we wish to reserve the word, "proposition", for something either true or false. A statement, in the form of a proposition, which fails to be either true or false, we regard as a mere group of symbols, without significance.

We may draw the analogy of a three dimensional geometry used in describing physical space, a case for which, we believe, the presence of such a situation is more commonly recognized. The entities of the geometry are clearly of abstract character, numbering as they do planes without thickness and points which cover no area in the plane, point sets containing an infinitude of points, lines of infinite length, and other things which cannot be reproduced in any physical experiment. Nevertheless the geometry can be applied to physical space in such a way that an extremely useful correspondence is set up between the theorems of the geometry and observable facts about material bodies in space. In building the geometry, the proposed application to physical space serves as a rough guide in determining what properties the abstract entities shall have, but does not assign these properties completely. Consequently there may be, and actually are, more than one geometry whose use is feasible in describing physical space. Similarly, there exist, undoubtedly, more than one formal system whose use as a logic is feasible, and of these systems one may be more pleasing or more convenient than another, but it cannot be said that one is right and the other wrong.

In consequence of this abstract character of the system which we are about to formulate, it is not admissible, in proving theorems of the system, to make use of the meaning of any of the symbols, although in the application which is intended the symbols do acquire meanings. The initial set of postulates must of themselves define the system as a formal structure, and in developing this formal structure reference to the proposed application must be held irrelevant. There may, indeed, be other applications of the system than its use as a logic.

4. **Intuitive logic.** It is clear, however, that formulas composed of symbols to which no meaning is attached cannot define a procedure of proof or justify an inference from one formula to another. If our postulates were expressed wholly by means of the symbols of formal logic without use of any words or symbols having a meaning, there would be no theorems except the postulates themselves. We are therefore obliged to use in some at least of our postulates other symbols than the undefined terms of the formal system, and to presuppose a knowledge of the meaning of these symbols, as well as to assume an understanding of a certain body of principles which these symbols are used to express, and which belong to what we shall call intuitive logic.[9] It seems desirable

[9] The principles of intuitive logic which we assume initially form, of course, a part of the body of facts to which the formal system, when completed, is to be applied. We should not, however, allow this to confuse us as to the clear cut distinction between intuitive logic and formal logic.

to make these presuppositions as few and as simple as we can, but there is no possibility of doing without them.

Before proceeding to the statement of our postulates, we shall attempt to make a list of these principles of intuitive logic which we find it necessary to assume and of the symbols a knowledge of whose meanings we presuppose. The latter belong to what we shall call the language of intuitive logic, as distinguished from the language of formal logic which is made up of the undefined terms of our abstract system.

We assume that we know the meaning of the words *symbol* and *formula* (by the word formula we mean a set of symbols arranged in an order of succession, one after the other). We assume the ability to write symbols and to arrange them in a certain order on a page, and the ability to recognize different occurrences of *the same* symbol and to distinguish between such a double occurrence of a symbol and the occurrence of *distinct* symbols. And we assume the possibility of dealing with a formula as a unit, of copying it at any desired point, and of recognizing other formulas as being *the same* or *distinct*.

We assume that we know what it means to say that a certain symbol or formula *occurs* in a given formula, and also that we are able to pick out and discuss a particular *occurrence* of one formula in another.

We assume an understanding of the operation of *substituting* a given symbol or formula *for a particular occurrence* of a given symbol or formula.

And we assume also an understanding of the operation of substitution throughout a given formula, and this operation we indicate by an S, $S^{\mathbf{X}}_{\mathbf{Y}}\mathbf{U}|$ representing the formula which results when we operate on the formula **U** by replacing **X** by **Y** throughout, where **Y** may be any symbol or formula but **X** must be a single symbol, not a combination of several symbols.

We assume that we know how to recognize a given formula as *being obtainable* from the formulas of a certain set by *repeated combinations* of the latter according to a given law. This assumption is used below in defining the term "well-formed". It may be described as an assumption of the ability to make a definition by induction, when dealing with groups of symbols.

We assume the ability to make the assertion that a given formula is one of those belonging to the abstract system which we are constructing, and this assertion we indicate by the words *is true*. As an abbreviation, however, we shall usually omit the words *is true*, the mere placing of the formula in an isolated position being taken as a sufficient indication of them.

We assume further the meaning and use of the word *every* as part of the language of intuitive logic, and the use in connection with it of variable letters, which we write in bold face type to distinguish them from variable letters used in the language of formal logic. These variable letters, written in bold face type, stand always for a variable (or undetermined) symbol or formula.

We assume the meaning and use of the following words from the language of intuitive logic: *there is, and, or, if · · · then, not,* and *is* in the sense of identity. That is, we assume that we know what combinations of these words with themselves and our other symbols constitute propositions, and, in a simple sense, what such propositions mean.

We assume that we know how to distinguish between the words and symbols we have been enumerating, which we shall describe as *symbols of intuitive logic,* and other symbols, which are mere symbols, without meaning, and which we shall describe as *formal symbols.*

We assume that we know what it is to be a proposition of intuitive logic, and that we are able to assert such propositions, not merely one proposition, but various propositions in succession. And, finally, we assume the permanency of a proposition once asserted, so that it may at any later stage be reverted to and used as if just asserted.

In making the preceding statements it becomes clear that a certain circle is unavoidable in that we are unable to make our explanations of the ideas in question intelligible to any but those who already understand at least a part of these ideas. For this reason we are compelled to assume them as known in the beginning independently of our statement of them. Our purpose has been, not to explain or convey these ideas, but to point out to those who already understand them what the ideas are to which we are referring and to explain our symbolism for them.

5. **Undefined terms.** We are now ready to set down a list of the undefined terms of our formal logic. They are as follows:

$$\{\ \}(\ ),\quad \lambda\ [\ ],\quad \Pi,\quad \Sigma,\quad \&,\quad \sim,\quad \iota,\quad A.$$

The expressions $\{\ \}(\ )$ and $\lambda\ [\ ]$ are not, of course, single symbols, but sets of several symbols, which, however, in every formula which will be provable as a consequence of our postulates, always occur in groups in the order here given, with other symbols or formulas between as indicated by the blank spaces.

In addition to the undefined terms just set down, we allow the use, in the formulas belonging to the system which we are constructing, of any other formal symbol, and these additional symbols used in our formulas we call *variables.*

An occurrence of a variable **x** in a given formula is called an occurrence of **x** as a *bound variable* in the given formula if it is an occurrence of **x** in a part of the formula of the form $\lambda \mathbf{x}[\mathbf{M}]$; that is, if there is a formula **M** such that $\lambda \mathbf{x}[\mathbf{M}]$ occurs in the given formula and the occurrence of **x** in question is an occurrence in $\lambda \mathbf{x}[\mathbf{M}]$. All other occurrences of a variable in a formula are called occurrences as a *free variable.*

A formula is said to be *well-formed* if it is a variable, or if it is one of the symbols Π, Σ, &, $\sim$, ι, A, or if it is obtainable from these symbols by repeated combinations of them of one of the forms $\{\mathbf{M}\}(\mathbf{N})$ and $\lambda \mathbf{x}[\mathbf{M}]$, where **x** is any variable and **M** and **N** are symbols or formulas which are being combined. This is a definition by induction. It implies the following rules: (1) a variable is well-formed (2) Π, Σ, &, $\sim$, ι, and A are well-formed (3) if **M** and **N** are well-formed then $\{\mathbf{M}\}(\mathbf{N})$ is well-formed (4) if **x** is a variable and **M** is well-formed then $\lambda \mathbf{x}[\mathbf{M}]$ is well-formed.

All the formulas which will be provable as consequences of our postulates will be well-formed and will contain no free variables.

The undefined terms of a formal system have, as we have explained, no meaning except in connection with a particular application of the system. But for the formal system which we are engaged in constructing we have in mind a particular application, which constitutes, in fact, the motive for constructing it, and we give here the meanings which our undefined terms are to have in this intended application.

If **F** is a function and **A** is a value of the independent variable for which the function is defined, then $\{\mathbf{F}\}(\mathbf{A})$ represents the value taken on by the function **F** when the independent variable takes on the value **A**. The usual notation is $\mathbf{F}(\mathbf{A})$. We introduce the braces on account of the possibility that **F** might be a combination of several symbols, but, in the case that **F** is a single symbol, we shall often use the notation $\mathbf{F}(\mathbf{A})$ as an abbreviation for the fuller expression.[10]

Adopting a device due to Schönfinkel,[11] we treat a function of two variables as a function of one variable whose values are functions of one variable, and a function of three or more variables similarly. Thus, what is usually written $\mathbf{F}(\mathbf{A},\mathbf{B})$ we write $\{\{\mathbf{F}\}(\mathbf{A})\}(\mathbf{B})$, and what is usually written $\mathbf{F}(\mathbf{A}, \mathbf{B}, \mathbf{C})$ we write $\{\{\{\mathbf{F}\}(\mathbf{A})\}(\mathbf{B})\}(\mathbf{C})$, and so on. But again we frequently find it convenient to employ the more usual notations as abbreviations.

If **M** is any formula containing the variable **x**, then $\lambda \mathbf{x}[\mathbf{M}]$ is a symbol for the function whose values are those given by the formula. That is,

[10] The braces { } are, as a matter of fact, superfluous and might have been omitted from our list of undefined terms, but their inclusion makes for readability of formulas.

[11] M. Schönfinkel, Über die Bausteine der mathematischen Logik, Math. Annalen, vol. 92 (1924), pp. 305–316.

$\lambda x[M]$ represents a function, whose value for a value L of the independent variable is equal to the result $S^x_L M|$ of substituting L for x throughout M, whenever $S^x_L M|$ turns out to have a meaning, and whose value is in any other case undefined.

The symbol Π stands for a certain propositional function of two independent variables, such that $\Pi(F, G)$ denotes, "$G(x)$ is a true proposition for all values of x for which $F(x)$ is a true proposition." It is necessary to distinguish between the proposition $\Pi(F, G)$ and the proposition $'x \cdot F(x) \supset G(x)$ (read, "For every x, $F(x)$ implies $G(x)$"). The latter proposition justifies, for any value M of x, the inference $F(M) \supset G(M)$, and hence can be used only in the case that the functions F and G are defined for all values of their respective independent variables. The proposition $\Pi(F, G)$ does not, on the other hand, justify this inference, although, when $\{F\}(M)$ is known to be true, it does justify the inference $\{G\}(M)$. And the proposition $\Pi(F, G)$ is, therefore, suitable for use in the case that the ranges of definition of the functions F and G are limited.

The symbol Σ stands for a certain propositional function of one independent variable, such that $\Sigma(F)$ denotes, "There exists at least one value of x for which $F(x)$ is true."

The symbol & stands for a certain propositional function of two independent variables, such that, if P and Q are propositions, &(P, Q) is the logical product P-and-Q.

The symbol $\sim$ stands for a certain propositional function of one independent variable, such that, if P is a proposition, then $\sim(P)$ is the negation of P and may be read, "Not-P".

The symbol ι stands for a certain function of one independent variable, such that, if F is a propositional function of one independent variable, then $\iota(F)$ denotes, "The object x such that $\{F\}(x)$ is true."

The symbol A stands for a certain function of two independent variables, the formula $A(F, M)$ being read, "The abstraction from M with respect to F."

6. **Abbreviations and definitions.** In practice we do not use actually the notation just described, but introduce various abbreviations and substituted notations, partly for the purpose of shortening our formulas and partly in order to render them more readable. We do not, however, regard these abbreviations as an essential part of our theory but rather as extraneous. When we use them we do not literally carry out the development of our system, but we do indicate in full detail how this development can be carried out, and this is for our purpose sufficient.

As has been said above, we use $\{F\}(A, B)$ as an abbreviation for $\{\{F\}(A)\}(B)$ and similarly in the case of functions of larger numbers of

variables. Moreover, alike in the case of functions of one variable and in the case of functions of two or more variables, we omit the braces { } whenever the function is represented by a single symbol rather than by an expression consisting of several symbols. Thus, if **F** is a single symbol, we write **F**(**A**) instead of {**F**}(**A**) and **F**(**A**, **B**) instead of {**F**}(**A**, **B**) or {{**F**}(**A**)}(**B**).

We shall usually write [**M**][**N**] instead of &(**M**, **N**) or {{&}(**M**)}(**N**), and ∼[**M**] instead of ∼(**M**) or {∼}(**M**).

When **F** is a single symbol, we shall often write **F**$\mathfrak{x}$[**M**] instead of **F**($\lambda\mathfrak{x}$[**M**]).

Instead of $\Pi(\lambda\mathfrak{x}[\mathbf{M}], \lambda\mathfrak{x}[\mathbf{N}])$, we shall write $[\mathbf{M}]\supset_{\mathfrak{x}}[\mathbf{N}]$. And $[\Sigma\mathfrak{y}[\mathbf{M}]]\supset_{\mathfrak{x}}[[\mathbf{M}]\supset_{\mathfrak{y}}[\mathbf{N}]]$ we abbreviate further to $[\mathbf{M}]\supset_{\mathfrak{xy}}[\mathbf{N}]$, and $[\Sigma\mathfrak{y}[\Sigma\mathfrak{z}[\mathbf{M}]]]\supset_{\mathfrak{x}}[[\Sigma\mathfrak{z}[\mathbf{M}]]\supset_{\mathfrak{y}}[[\mathbf{M}]\supset_{\mathfrak{z}}[\mathbf{N}]]]$ we abbreviate to $[\mathbf{M}]\supset_{\mathfrak{xyz}}[\mathbf{N}]$ and so on.

Moreover, whenever possible without ambiguity, we omit square brackets [ ], whether the brackets belong to the undefined term $\lambda[\ ]$ or whether they appear as a part of one of our abbreviations. In order to allow the omission of square brackets as often as possible, we adopt the convention that, whenever there are more than one possibility, the extent of the omitted square brackets shall be taken as the shortest. And when the omission of the square brackets is not possible without ambiguity, we can sometimes substitute for them a dot, or period. This dot, when it occurs within a parenthesis, enclosed by either square brackets [ ], round parentheses ( ), or braces { }, stands for square brackets extending from the place where the dot occurs and up to the end of the parenthesis, or, if the parenthesis is divided into sections by commas as in the case of functions of two or more variables, extending from the place where the dot occurs and up to the first of these commas or to the end of the parenthesis, whichever is first reached. And when the dot is not within any parenthesis, it stands for square brackets extending from the place where it occurs and up to the end of the entire formula. In other words, a dot represents square brackets extending the greatest possible distance forward from the point where it occurs.

In addition to these abbreviations, we allow afreely the introduction of abbreviations of a simpler sort, which we call definitions,[12] and which consist in the substitution of a particular single symbol for a particular well-formed formula containing no free variables.

[12] There seems to be, as a matter of fact, no serious objection to treating definitions as an essential part of the system rather than as extraneous, but we believe it more consistent to class them with our other and more complicated abbreviations.

We introduce at once the following definitions, using an arrow $\longrightarrow$ to mean "Stands for", or, "Is an abbreviation for":

$V \longrightarrow \lambda\mu\lambda\nu \,.\sim.\sim\mu\,.\sim\nu$

$U \longrightarrow \lambda\mu\lambda\nu \,.\sim.\,\mu\,.\sim\nu$

$Q \longrightarrow \lambda\mu\lambda\nu \,.\, \Pi(\mu, \nu)\,.\,\Pi(\nu, \mu)$

$E \longrightarrow \lambda\pi\Sigma\varphi \,.\, \varphi(\pi)$

$V(\mathbf{P}, \mathbf{Q})$ is to be read, "**P** or **Q**", $U(\mathbf{P}, \mathbf{Q})$ is to be read, "**P** implies **Q**", $Q(\mathbf{F}, \mathbf{G})$ is to be read "**F** and **G** are equivalent" and $E(\mathbf{M})$ is to be read "**M** exists"

And in connection with these symbols just defined we introduce some further abbreviations. For $V(\mathbf{P}, \mathbf{Q})$ we write $[\mathbf{P}]\,\mathrm{v}\,[\mathbf{Q}]$, and for $U(P, Q)$ we write $[\mathbf{P}]\supset[\mathbf{Q}]$. We abbreviate $Q(\lambda\mathbf{x}\,.\,\mathbf{M},\ \lambda\mathbf{x}\,.\,\mathbf{N})$ to $[\mathbf{M}] \equiv_{\mathbf{x}} [\mathbf{N}]$. And we abbreviate $E(\mathbf{x})\,\supset_{\mathbf{x}}[\mathbf{M}]$ to $\mathrm{'}\mathbf{x}[\mathbf{M}]$, which may be read, "For every **x**, **M**".

The notion of a class may be introduced by means of the definition:

$$K \longrightarrow A(Q).$$

The formula $K(\mathbf{F})$ is then to be read, "the class of x's such that $\{\mathbf{F}\}(x)$ true."

7. **Postulates.** We divide our postulates into two groups, of which the first consists of what we shall call rules of procedure and the second of what we shall call formal postulates. The latter assert that a given formula is true, and contain nothing from the language of intuitive logic other than the words *is true* (and even these words, as already explained, we leave unexpressed when we write the postulates). And the former, the rules of procedure, contain other words from the language of intuitive logic.

The theorems which are proved as consequences of these postulates are of the same form as the postulates of the first group, namely, that a certain formula is true. And the proof of a theorem consists of a series of steps which, from a set of one or more postulates of the first group as a starting point, leads us to the theorem, each step being justified by an appeal to one of the rules of procedure.

The postulates of our first group, the rules of procedure, are five in number:

I. *If* **J** *is true, if* **L** *is well-formed, if all the occurrences of the variable* **x** *in* **L** *are occurrences as a bound variable, and if the variable* **y** *does not occur in* **L**, *then* **K**, *the result of substituting* $\mathrm{S}^{\mathbf{x}}_{\mathbf{y}}\mathbf{L}|$ *for a particular occurrence of* **L** *in* **J**, *is also true.*

II. *If* **J** *is true, if* **M** *and* **N** *are well-formed, if the variable* **x** *occurs in* **M**, *and if the bound variables in* **M** *are distinct both from the variable* **x** *and from the free variables in* **N**, *then* **K**, *the result of substituting* $\mathrm{S}^{\mathbf{x}}_{\mathbf{N}}\mathbf{M}|$ *for a particular occurrence of* $\{\lambda\mathbf{x}\,.\,\mathbf{M}\}(\mathbf{N})$ *in* **J**, *is also true.*

III. *If* $\mathbf{J}$ *is true, if* $\mathbf{M}$ *and* $\mathbf{N}$ *are well-formed, if the variable* $\mathbf{x}$ *occurs in* $\mathbf{M}$, *and if the bound variables in* $\mathbf{M}$ *are distinct both from* $\mathbf{x}$ *and from the free variables in* $\mathbf{N}$, *then* $\mathbf{K}$, *the result of substituting* $\{\lambda\mathbf{x}\,.\,\mathbf{M}\}(\mathbf{N})$ *for a particular occurrence of* $S^{\mathbf{x}}_{\mathbf{N}}\mathbf{M}|$ *in* $\mathbf{J}$, *is also true.*

IV. *If* $\{\mathbf{F}\}(\mathbf{A})$ *is true and* $\mathbf{F}$ *and* $\mathbf{A}$ *are well-formed, then* $\Sigma(\mathbf{F})$ *is true.*

V. *If* $\Pi(\mathbf{F}, \mathbf{G})$ *and* $\{\mathbf{F}\}(\mathbf{A})$ *are true, and* $\mathbf{F}$, $\mathbf{G}$, *and* $\mathbf{A}$ *are well-formed, then* $\{\mathbf{G}\}(\mathbf{A})$ *is true.*

And our formal postulates are the thirty-seven following:

1. $\Sigma(\varphi) \supset_{\varphi} \Pi(\varphi, \varphi)$.
2. $'x\,.\,\varphi(x) \supset_{\varphi} .\, \Pi(\varphi, \psi) \supset_{\psi} \psi(x)$.
3. $\Sigma(\sigma) \supset_{\sigma} .\, [\sigma(x) \supset_x \varphi(x)] \supset_{\varphi} .\, \Pi(\varphi, \psi) \supset_{\psi} .\, \sigma(x) \supset_x \psi(x)$.
4. $\Sigma(\varrho) \supset_{\varrho} .\, \Sigma y\,[\varrho(x) \supset_x \varphi(x, y)] \supset_{\varphi} .\, [\varrho(x) \supset_x \Pi(\varphi(x), \psi(x))] \supset_{\psi} .$ $[\varrho(x) \supset_x \varphi(x, y)] \supset_y .\, \varrho(x) \supset_x \psi(x, y)$.
5. $\Sigma(\varphi) \supset_{\varphi} .\, \Pi(\varphi, \psi) \supset_{\psi} .\, \varphi(f(x)) \supset_{fx} \psi(f(x))$.
6. $'x\,.\,\varphi(x) \supset_{\varphi} .\, \Pi(\varphi, \psi(x)) \supset_{\psi} \psi(x, x)$.
7. $\varphi(x, f(x)) \supset_{\varphi fx} .\, \Pi(\varphi(x), \psi(x)) \supset_{\psi} \psi(x, f(x))$.
8. $\Sigma(\varrho) \supset_{\varrho} .\, \Sigma y\,[\varrho(x) \supset_x \varphi(x, y)] \supset_{\varphi} .\, [\varrho(x) \supset_x \Pi(\varphi(x), \psi)] \supset_{\psi} .$ $[\varrho(x) \supset_x \varphi(x, y)] \supset_y \psi(y)$.
9. $'x\,.\,\varphi(x) \supset_{\varphi} \Sigma(\varphi)$.
10. $\Sigma x\,\varphi(f(x)) \supset_{f\varphi} \Sigma(\varphi)$.
11. $\varphi(x, x) \supset_{\varphi x} \Sigma(\varphi(x))$.
12. $\Sigma(\varphi) \supset_{\varphi} \Sigma x\,\varphi(x)$.
13. $\Sigma(\varphi) \supset_{\varphi} .\, [\varphi(x) \supset_x \psi(x)] \supset_{\psi} \Pi(\varphi, \psi)$.
14. $p \supset_p .\, q \supset_q p\,q$.
15. $p\,q \supset_{qp} p$.
16. $p\,q \supset_{pq} q$.
17. $\Sigma x\,\Sigma\theta\,[\varphi(x) .\sim \theta(x) .\, \Pi(\psi, \theta)] \supset_{\varphi\psi} \sim \Pi(\varphi, \psi)$.
18. $\sim \Pi(\varphi, \psi) \supset_{\varphi\psi} \Sigma x\,\Sigma\theta\,.\,\varphi(x) .\sim \theta(x) .\, \Pi(\psi, \theta)$.
19. $\Sigma x\,\Sigma\theta\,[\sim \varphi(u, x) .\sim \theta(u) .\, \Sigma(\varphi(y)) \supset_y \theta(y)] \supset_{\varphi u} \sim \Sigma(\varphi(u))$.
20. $\sim \Sigma(\varphi) \supset_{\varphi} \Sigma x\,.\sim \varphi(x)$.
21. $p \supset_p .\sim q \supset_q \sim .\, p\,q$.
22. $\sim p \supset_p .\, q \supset_q \sim .\, p\,q$.
23. $\sim p \supset_p .\sim q \supset_q \sim .\, p\,q$.
24. $p \supset_p .\, [\sim .\, p\,q] \supset_q \sim q$.

25. $[\sim . \varphi(u) . \psi(u)] \supset_{\varphi\psi u} . [[\varphi(x) . \sim \psi(x)] \supset_x \varrho(x)] .$
$[[\sim \varphi(x) . \psi(x)] \supset_x \varrho(x)] . [\sim \varphi(x) . \sim \psi(x)] \supset_x \varrho(x)] \supset_\varrho \varrho(u).$

26. $p \supset_p \sim \sim p.$

27. $\sim \sim p \supset_p p.$

28. $\sim \Sigma(\varphi) \supset_\varphi . \Sigma(\psi) \supset_\psi \Pi(\varphi, \psi).$

29. $\sim \Sigma(\varphi) \supset_\varphi . \sim \Sigma(\psi) \supset_\psi \Pi(\varphi, \psi).$

30. $\iota x . \varphi(x) \supset_\varphi . [\theta(x) . \psi(x) \supset_\psi \Pi(\theta, \psi)] \supset_\theta \varphi(\iota(\theta)).$

31. $\varphi(\iota(\theta)) \supset_{\theta\varphi} \Pi(\theta, \varphi).$

32. $E(\iota(\theta)) \supset_\theta \Sigma(\theta).$

33. $[\varphi(x, y) \supset_{xy} \varphi(y, z) \supset_z \varphi(x, z)] \ [\varphi(x, y) \supset_{xy} \varphi(y, x)] \supset_\varphi . \varphi(u, v) \supset_{uv} .$
$\psi(A(\varphi, u)) \supset_\psi \psi(A(\varphi, v)).$

34. $[\varphi(x, y) \supset_{xy} . \varphi(y, z) \supset_z \varphi(x, z)] \ [\varphi(x, y) \supset_{xy} \varphi(y, x)] \supset_\varphi .$
$[\varphi(x, y) \supset_{xy} \theta(x, y)] \supset_\theta . \sim \theta(u, v) \supset_{uv} \sim . \psi(A(\varphi, u)) \supset_\psi \psi(A(\varphi, v)).$

35. $[\psi(A(\varphi, u)) \supset_\psi \psi(A(\varphi, v))] \supset_{\varphi uv} \varphi(u, v).$

36. $E(A(\varphi)) \supset_\varphi . \varphi(x, y) \supset_{xy} . \varphi(y, z) \supset_z \varphi(x, z).$

37. $E(A(\varphi)) \supset_\varphi . \varphi(x, y) \supset_{xy} \varphi(y, x).$

8. **The relation between free and bound variables.** By a *step* in a proof we mean an application of one of the rules of procedure IV or V, occurring in the course of the proof. And in counting the *number of steps* in a proof, each step is to be counted with its proper multiplicity. That is, if a formula **M** is proved and then used r times as a premise for subsequent steps of the proof, then each step in the proof of **M** is to be counted r times.

If **M** and **N** are well-formed and if **N** can be derived from **M** by successive applications of the rules of procedure I, II, and III, then **M** is said to be *convertible* into **N**, and the process is spoken of as a *conversion* of **M** into **N**.

The formula **N** is said to be *provable as a consequence of* the formula **M**, if **M** is well-formed, and **N** could be made a provable formula by adding **M** to our list of formal postulates as a thirty-eighth postulate. Either of the formulas, **M** or **N**, or both, may contain free variables, since although none of our formal postulates contains free variables, there is, formally, nothing to prevent our adding a thirty-eighth postulate which does contain free variables.

We conclude by proving about our system of postulates the three following theorems:

THEOREM I. *Suppose that* $\mathbf{M}$ *contains* $\mathbf{x}$ *as a free variable, and that* $\Sigma\mathbf{x}.\mathbf{M}$ *is provable, and that* $\mathbf{N}$ *is provable as a consequence of* $\mathbf{M}$. *Then, if* $\mathbf{N}$ *contains* $\mathbf{x}$ *as a free variable, the formula* $[\mathbf{M}]\supset_{\mathbf{x}}.\mathbf{N}$ *is provable. And if* $\mathbf{N}$ *does not contain* $\mathbf{x}$ *as a free variable, the formula* $\mathbf{N}$ *itself is provable.*

We shall prove this theorem by induction with respect to the number of steps in the proof of $\mathbf{N}$ as a consequence of $\mathbf{M}$.

If this number of steps is zero, then $\mathbf{M}$ is convertible into $\mathbf{N}$, and, since $\mathbf{M}$ contains $\mathbf{x}$ as a free variable, so does $\mathbf{N}$. Hence in this case $[\mathbf{M}]\supset_{\mathbf{x}}.\mathbf{N}$ may be proved as follows.

Before each formula we give a symbol which will subsequently be used in referring to it. And after each formula we give the means by which it is inferred, that is, the number of the rule of procedure and a reference to the premise or premises, or, if the formula is one of our formal postulates, the number of that postulate.

$\mathbf{B}_1$: $\Sigma\mathbf{x}.\mathbf{M}$ —provable by hypothesis.

$\mathbf{A}_1$: $\{\lambda\varphi\,\Sigma(\varphi)\}\,(\lambda\mathbf{x}.\mathbf{M})$ —III, $\mathbf{B}_1$.

$\mathbf{A}_2$: $\Sigma(\varphi)\supset_{\varphi}\Pi(\varphi,\varphi)$ —1.

$\mathbf{A}_3$: $\{\lambda\varphi\,\Pi(\varphi,\varphi)\}\,(\lambda\mathbf{x}.\mathbf{M})$ —V, $\mathbf{A}_2$, $\mathbf{A}_1$.

$\mathbf{A}_4$: $[\mathbf{M}]\supset_{\mathbf{x}}.\mathbf{M}$ —II, $\mathbf{A}_3$.

$\mathbf{A}_5$: $[\mathbf{M}]\supset_{\mathbf{x}}.\mathbf{N}$ —by conversion, from $\mathbf{A}_4$.

Suppose now, as the hypothesis of our induction, that our theorem is true whenever the number of steps in the proof of $\mathbf{N}$ as a consequence of $\mathbf{M}$ is not greater than n. And consider a case in which this number of steps is $n+1$.

The last step in the proof of $\mathbf{N}$ as a consequence of $\mathbf{M}$ might be an application of Rule IV, with premise $\{\mathbf{F}\}(\mathbf{A})$ and conclusion $\Sigma(\mathbf{F})$ or it might be an application of Rule V, with premises $\Pi(\mathbf{F},\mathbf{G})$ and $\{\mathbf{F}\}(\mathbf{A})$ and conclusion $\{\mathbf{G}\}(\mathbf{A})$. In the former case, $\mathbf{F}$ might or might not contain $\mathbf{x}$ as a free variable, and $\mathbf{A}$ might or might not contain $\mathbf{x}$ as a free variable. In the latter case, $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{A}$ each might or might not contain $\mathbf{x}$ as a free variable. All these possibilities must be considered separately.

Case 1, the last step in the proof of $\mathbf{N}$ as a consequence of $\mathbf{M}$ is an application of Rule V, $\mathbf{A}$ is identical with $\mathbf{x}$, and neither $\mathbf{F}$ nor $\mathbf{G}$ contains $\mathbf{x}$ as a free variable. We can prove $[\mathbf{M}]\supset_{\mathbf{x}}.\mathbf{N}$ as follows:

$\mathbf{B}_1$: $\Sigma\mathbf{x}.\mathbf{M}$ —provable, by hypothesis.

$\mathbf{A}_1$: $\{\lambda(\sigma)\,\Sigma(\sigma)\}\,(\lambda\mathbf{x}.\mathbf{M})$ —III, $\mathbf{B}_1$.

$\mathbf{A}_2$: $\Sigma(\sigma)\supset_{\sigma}.[\sigma(x)\supset_x\varphi(x)]\supset_{\varphi}.\Pi(\varphi,\psi)\supset_{\psi}.\sigma(x)\supset_x\psi(x)$ —3.

$\mathbf{A}_3$: $\{\lambda\sigma.[\sigma(x)\supset_x\varphi(x)]\supset_{\varphi}.\Pi(\varphi,\psi)\supset_{\psi}.\sigma(x)\supset_x\psi(x)\}\,(\lambda\mathbf{x}.\mathbf{M})$ —V, $\mathbf{A}_2$, $\mathbf{A}_1$.

$\mathbf{A}_4$: $\{\lambda\sigma \cdot [\sigma(\mathbf{x}) \supset_{\mathbf{x}} \varphi(\mathbf{x})] \supset_{\varphi} \cdot \Pi(\varphi, \psi) \supset_{\psi} \cdot \sigma(\mathbf{x}) \supset_{\mathbf{x}} \psi(\mathbf{x})\}(\lambda\mathbf{x} \cdot \mathbf{M})$ —I, $\mathbf{A}_3$.

$\mathbf{A}_5$: $[\{\lambda\mathbf{x} \cdot \mathbf{M}\}(\mathbf{x}) \supset_{\mathbf{x}} \varphi(\mathbf{x})] \supset_{\varphi} \cdot \Pi(\varphi, \psi) \supset_{\psi} \cdot \{\lambda\mathbf{x} \cdot \mathbf{M}\}(\mathbf{x}) \supset_{\mathbf{x}} \psi(\mathbf{x})$ —II, $\mathbf{A}_4$.

$\mathbf{A}_6$: $[[\mathbf{M}] \supset_{\mathbf{x}} \varphi(\mathbf{x})] \supset_{\varphi} \cdot \Pi(\varphi, \psi) \supset_{\psi} \cdot \{\lambda\mathbf{x} \cdot \mathbf{M}\}(\mathbf{x}) \supset_{\mathbf{x}} \psi(\mathbf{x})$ —II, $\mathbf{A}_5$.[13]

$\mathbf{A}_7$: $[[\mathbf{M}] \supset_{\mathbf{x}} \varphi(\mathbf{x})] \supset_{\varphi} \cdot \Pi(\varphi, \psi) \supset_{\psi} \cdot [\mathbf{M}] \supset_{\mathbf{x}} \psi(\mathbf{x})$ —II, $\mathbf{A}_6$.

$\mathbf{B}_2$: $[\mathbf{M}] \supset_{\mathbf{x}} \{\mathbf{F}\}(\mathbf{x})$ —provable, by hypothesis of induction.

$\mathbf{A}_8$: $\{\lambda\varphi \cdot [\mathbf{M}] \supset_{\mathbf{x}} \varphi(\mathbf{x})\}(\mathbf{F})$ —III, $\mathbf{B}_2$.

$\mathbf{A}_9$: $\{\lambda\varphi \cdot \Pi(\varphi, \psi) \supset_{\psi} \cdot [\mathbf{M}] \supset_{\mathbf{x}} \psi(\mathbf{x})\}(\mathbf{F})$ —V, $\mathbf{A}_7$, $\mathbf{A}_8$.

$\mathbf{A}_{10}$: $\Pi(\mathbf{F}, \psi) \supset_{\psi} \cdot [\mathbf{M}] \supset_{\mathbf{x}} \psi(\mathbf{x})$ —II, $\mathbf{A}_9$.

$\mathbf{B}_3$: $\Pi(\mathbf{F}, \mathbf{G})$ —provable, by hypothesis of induction.

$\mathbf{A}_{11}$: $\{\lambda\psi\,\Pi(\mathbf{F}, \psi)\}(\mathbf{G})$ —III, $\mathbf{B}_3$.

$\mathbf{A}_{12}$: $\{\lambda\psi \cdot [\mathbf{M}] \supset_{\mathbf{x}} \psi(\mathbf{x})\}(\mathbf{G})$ —V, $\mathbf{A}_{10}$, $\mathbf{A}_{11}$.

$\mathbf{A}_{13}$: $[\mathbf{M}] \supset_{\mathbf{x}} \{\mathbf{G}\}(\mathbf{x})$ —II, $\mathbf{A}_{12}$.

$\mathbf{A}_{14}$: $[\mathbf{M}] \supset_{\mathbf{x}} \cdot \mathbf{N}$ —by conversion, from $\mathbf{A}_{13}$, since $\{\mathbf{G}\}(\mathbf{x})$ is convertiable into $\mathbf{N}$.

Case 2, the last step in the proof of $\mathbf{N}$ as a consequence of $\mathbf{M}$ is an application of Rule V, and $\mathbf{F}$ contains $\mathbf{x}$ as a free variable, and $\mathbf{G}$ and $\mathbf{A}$ do not. We can prove $\mathbf{N}$ as follows:

$\mathbf{B}_1$: $\Sigma\mathbf{x} \cdot \mathbf{M}$ —provable by hypothesis.

$\mathbf{A}_1$: $\{\lambda\varrho\,\Sigma(\varrho)\}(\lambda\mathbf{x} \cdot \mathbf{M})$ —III, $\mathbf{B}_1$.

$\mathbf{A}_2$: $\Sigma(\varrho) \supset_{\varrho} \cdot \Sigma y\,[\varrho(x) \supset_x \varphi(x, y)] \supset_{\varphi} \cdot [\varrho(x) \supset_x \Pi(\varphi(x), \psi)] \supset_{\psi} \cdot$
$[\varrho(x) \supset_x \varphi(x, y)] \supset_y \psi(y)$ —8.

$\mathbf{A}_3$: $\{\lambda\varrho \cdot \Sigma y\,[\varrho(x) \supset_x \varphi(x, y)] \supset_{\varphi} \cdot [\varrho(x) \supset_x \Pi(\varphi(x), \psi)] \supset_{\psi} \cdot$
$[\varrho(x) \supset_x \varphi(x, y)] \supset_y \psi(y)\}(\lambda\mathbf{x} \cdot \mathbf{M})$ —V, $\mathbf{A}_2$, $\mathbf{A}_1$.

$\mathbf{A}_4$: $\{\lambda\varrho \cdot \Sigma y\,[\varrho(\mathbf{x}) \supset_{\mathbf{x}} \varphi(\mathbf{x}, y)] \supset_{\varphi} \cdot [\varrho(\mathbf{x}) \supset_{\mathbf{x}} \Pi(\varphi(\mathbf{x}), \psi)] \supset_{\psi} \cdot$
$[\varrho(\mathbf{x}) \supset_{\mathbf{x}} \varphi(\mathbf{x}, y)] \supset_y \psi(y)\}(\lambda\mathbf{x} \cdot \mathbf{M})$ —I, $\mathbf{A}_3$.

$\mathbf{A}_5$: $\Sigma y\,[\{\lambda\mathbf{x} \cdot \mathbf{M}\}(\mathbf{x}) \supset_{\mathbf{x}} \varphi(\mathbf{x}, y)] \supset_{\varphi} \cdot [\{\lambda\mathbf{x} \cdot \mathbf{M}\}(\mathbf{x}) \supset_{\mathbf{x}} \Pi(\varphi(\mathbf{x}), \psi)] \supset_{\psi} \cdot$
$[\{\lambda\mathbf{x} \cdot \mathbf{M}\}(\mathbf{x}) \supset_{\mathbf{x}} \varphi(\mathbf{x}, y)] \supset_y \psi(y)$ —II, $\mathbf{A}_4$.

$\mathbf{A}_6$: $\Sigma y\,[[\mathbf{M}] \supset_{\mathbf{x}} \varphi(\mathbf{x}, y)] \supset_{\varphi} \cdot [\{\lambda\mathbf{x} \cdot \mathbf{M}\}(\mathbf{x}) \supset_{\mathbf{x}} \Pi(\varphi(\mathbf{x}), \psi)] \supset_{\psi} \cdot$
$[\{\lambda\mathbf{x} \cdot \mathbf{M}\}(\mathbf{x}) \supset_{\mathbf{x}} \varphi(\mathbf{x}, y)] \supset_y \psi(y)$ —II, $\mathbf{A}_5$.

$\mathbf{A}_7$: $\Sigma y\,[[\mathbf{M}] \supset_{\mathbf{x}} \varphi(\mathbf{x}, y)] \supset_{\varphi} \cdot [[\mathbf{M}] \supset_{\mathbf{x}} \Pi(\varphi(\mathbf{x}), \psi)] \supset_{\psi} \cdot$
$[\{\lambda\mathbf{x} \cdot \mathbf{M}\}(\mathbf{x}) \supset_{\mathbf{x}} \varphi(\mathbf{x}, y)] \supset_y \psi(y)$ —II, $\mathbf{A}_6$.

$\mathbf{A}_8$: $\Sigma y\,[[\mathbf{M}] \supset_{\mathbf{x}} \varphi(\mathbf{x}, y)] \supset_{\varphi} \cdot [[\mathbf{M}] \supset_{\mathbf{x}} \Pi(\varphi(\mathbf{x}), \psi)] \supset_{\psi} \cdot$
$[[\mathbf{M}] \supset_{\mathbf{x}} \psi(\mathbf{x}, y)] \supset_y \psi(y)$ —II, $\mathbf{A}_7$.

[13] Evidently, we may suppose, without loss of generality, that $\mathbf{M}$ does not contain $\mathbf{x}$ as a bound variable, because, if the formula $\mathbf{M}$ did contain $\mathbf{x}$ as a bound variable, it would be convertible into a formula $\mathbf{M}'$ which did not contain $\mathbf{x}$ as a bound variable. A similar remark applies frequently below.

$\mathbf{B}_2$: $[\mathbf{M}]\supset_{\mathbf{x}}\{\mathbf{F}\}(\mathbf{A})$ —provable, by hypothesis of induction.

$\mathbf{A}_9$: $[\mathbf{M}]\supset_{\mathbf{x}}\{\lambda\mathbf{x}\,.\,\mathbf{F}\}(\mathbf{x},\mathbf{A})$ —III, $\mathbf{B}_2$.

$\mathbf{A}_{10}$: $\{\lambda y\,.\,[\mathbf{M}]\supset_{\mathbf{x}}\{\lambda\mathbf{x}\,.\,\mathbf{F}\}(\mathbf{x},y)\}(\mathbf{A})$ —III, $\mathbf{A}_9$.

$\mathbf{A}_{11}$: $\Sigma y\,.\,[\mathbf{M}]\supset_{\mathbf{x}}\{\lambda\mathbf{x}\,.\,\mathbf{F}\}(\mathbf{x},y)$ —IV, $\mathbf{A}_{10}$.

$\mathbf{A}_{12}$: $\{\lambda\varphi\,\Sigma y\,.\,[\mathbf{M}]\supset_{\mathbf{x}}\varphi(\mathbf{x},y)\}(\lambda\mathbf{x}\,.\,\mathbf{F})$ —III, $\mathbf{A}_{11}$.

$\mathbf{A}_{13}$: $\{\lambda\varphi\,.\,[[\mathbf{M}]\supset_{\mathbf{x}}\Pi(\varphi(\mathbf{x}),\psi)]\supset_{\psi}\cdot$
$[\mathbf{M}]\supset_{\mathbf{x}}\varphi(\mathbf{x},y)]\supset_y\psi(y)\}(\lambda\mathbf{x}\,.\,\mathbf{F})$ —V, $\mathbf{A}_8$, $\mathbf{A}_{12}$.

$\mathbf{A}_{14}$: $[[\mathbf{M}]\supset_{\mathbf{x}}\Pi(\{\lambda\mathbf{x}\,.\,\mathbf{F}\}(\mathbf{x}),\psi)]\supset_{\psi}\cdot[[\mathbf{M}]\supset_{\mathbf{x}}\{\lambda\mathbf{x}\,.\,\mathbf{F}\}(\mathbf{x},y)]\supset_y\psi(y)$ —II, $\mathbf{A}_{13}$.

$\mathbf{A}_{15}$: $[[\mathbf{M}]\supset_{\mathbf{x}}\Pi(\mathbf{F},\psi)]\supset_{\psi}\cdot[[\mathbf{M}]\supset_{\mathbf{x}}\cdot\{\lambda\mathbf{x}\,.\,\mathbf{F}\}(\mathbf{x},y)]\supset_y\psi(y)$ —II, $\mathbf{A}_{14}$.

$\mathbf{A}_{16}$: $[[\mathbf{M}]\supset_{\mathbf{x}}\Pi(\mathbf{F},\psi)]\supset_{\psi}\cdot[[\mathbf{M}]\supset_{\mathbf{x}}\{\mathbf{F}\}(y)]\supset_y\psi(y)$ —II, $\mathbf{A}_{15}$.

$\mathbf{B}_3$: $[\mathbf{M}]\supset_{\mathbf{x}}\Pi(\mathbf{F},\mathbf{G})$ —provable, by hypothesis of induction.

$\mathbf{A}_{17}$: $\{\lambda\psi\,.\,[\mathbf{M}]\supset_{\mathbf{x}}\Pi(\mathbf{F},\psi)\}(\mathbf{G})$ —III, $\mathbf{B}_3$.

$\mathbf{A}_{18}$: $\{\lambda\psi\,.\,[[\mathbf{M}]\supset_{\mathbf{x}}\{\mathbf{F}\}(y)]\supset_y\psi(y)\}(\mathbf{G})$ —V, $\mathbf{A}_{16}$, $\mathbf{A}_{17}$.

$\mathbf{A}_{19}$: $[[\mathbf{M}]\supset_{\mathbf{x}}\{\mathbf{F}\}(y)]\supset_{\mathbf{x}}\{\mathbf{G}\}(y)$ —II, $\mathbf{A}_{18}$.

$\mathbf{A}_{20}$: $\{\lambda y\,.\,[\mathbf{M}]\supset_{\mathbf{x}}\{\mathbf{F}\}(y)\}(\mathbf{A})$ —III, $\mathbf{B}_2$.

$\mathbf{A}_{21}$: $\{\lambda y\{\mathbf{G}\}(y)\}(\mathbf{A})$ —V, $\mathbf{A}_{19}$, $\mathbf{A}_{20}$.

$\mathbf{A}_{22}$: $\{\mathbf{G}\}(\mathbf{A})$ —II, $\mathbf{A}_{21}$.

$\mathbf{N}$ —by conversion, from $\mathbf{A}_{22}$.

Case 3, the last step in the proof of $\mathbf{N}$ as a consequence of $\mathbf{M}$ is an application of Rule IV, $\mathbf{A}$ is identical with $\mathbf{x}$, and $\mathbf{F}$ does not contain $\mathbf{x}$ as a free variable. By hypothesis, $\Sigma\mathbf{x}\,.\,\mathbf{M}$ is provable, and from it, by conversion and an application of Rule IV, we can obtain $\Sigma\varphi\,.\,\varphi(\lambda\mathbf{x}\,.\,\mathbf{M})$. Therefore, using Postulate 9, we can prove $\varphi(\lambda\mathbf{x}\,.\,\mathbf{M})\supset_{\varphi}\Sigma(\varphi)$. And hence, using Postulate 5, we can prove $f(\mathbf{x},\lambda\mathbf{x}\,.\,\mathbf{M})\supset_{f\mathbf{x}}\Sigma(f(\mathbf{x}))$. Moreover $\Sigma\mathbf{x}\,.\,\mathbf{M}$ becomes by conversion $\Sigma\mathbf{x}\,.\,\{\lambda y\,\lambda\varphi\,.\,\varphi(y)\}(\mathbf{x},\lambda\mathbf{x}\,.\,\mathbf{M})$, and this combined with the preceding yields $\{\lambda y\,\lambda\varphi\,.\,\varphi(y)\}(\mathbf{x},\lambda\mathbf{x}\,.\,\mathbf{M})\supset_{\mathbf{x}}\Sigma(\{\lambda y\,\lambda\varphi\,.\,\varphi(y)\}(\mathbf{x}))$. And this last formula becomes by conversion $[\mathbf{M}]\supset_{\mathbf{x}}\{\lambda\mathbf{x}\,E(\mathbf{x})\}(\mathbf{x})$.

Now using Postulate 3, since $\Sigma\mathbf{x}\,.\,\mathbf{M}$ is provable, we can obtain $[[\mathbf{M}]\supset_{\mathbf{x}}\varphi(\mathbf{x})]\supset_{\varphi}\cdot\Pi(\varphi,\psi)\supset_{\psi}\cdot[\mathbf{M}]\supset_{\mathbf{x}}\psi(\mathbf{x})$. Hence, using the last formula in the preceding paragraph, we can prove $\Pi(\lambda\mathbf{x}E(\mathbf{x}),\psi)\supset_{\psi}\cdot[\mathbf{M}]\supset_{\mathbf{x}}\psi(\mathbf{x})$. And hence, using Postulate 9, we can prove $[\mathbf{M}]\supset_{\mathbf{x}}\cdot\varphi(\mathbf{x})\supset_{\varphi}\Sigma(\varphi)$.

Thus we are able to prove $[\mathbf{M}]\supset_{\mathbf{x}}\Pi(\mathbf{F}',\mathbf{G}')$, where $\mathbf{F}'$ stands for $\lambda\varphi\,.\,\varphi(x)$ and $\mathbf{G}'$ for $\lambda\varphi\,\Sigma(\varphi)$. And, by the hypothesis of our induction, we are able to prove $[\mathbf{M}]\supset_{\mathbf{x}}\{\mathbf{F}'\}(\mathbf{A}')$, where $\mathbf{A}'$ stands for $\mathbf{F}$. Therefore, by the method of Case 2, we can prove $\{\mathbf{G}'\}(\mathbf{A}')$, and from it by conversion, $\mathbf{N}$.

Case 4, the last step in the proof of $\mathbf{N}$ as a consequence of $\mathbf{M}$ is an application of Rule IV, and $\mathbf{A}$ contains $\mathbf{x}$ as a free variable, and $\mathbf{F}$ does not. Since $\{\mathbf{F}\}(\mathbf{A})$ is convertible into $\{\lambda\mathbf{x}\{\mathbf{F}\}(\mathbf{A})\}(\mathbf{x})$, we are able, by the

method of Case 3, to prove $\Sigma \mathbf{x} . \{\mathbf{F}\}(\mathbf{A})$. This last formula is convertible into $\Sigma \mathbf{x} . \{\mathbf{F}\}(\{\lambda \mathbf{x} . \mathbf{A}\}(\mathbf{x}))$. Hence, using Postulate 10, we can prove $\Sigma(\mathbf{F})$ and from it by conversion, **N**.

Case 5, the last step in the proof of **N** as a consequence of **M** is an application of Rule IV, and **F** contains **x** as a free variable, and **A** does not. By the hypothesis of our induction, we can prove $[\mathbf{M}] \supset_{\mathbf{x}} \{\mathbf{F}\}(\mathbf{A})$. Hence we can prove $E(\mathbf{A})$, and hence, using Postulate 9, $\varphi(\mathbf{A}) \supset_{\varphi} \Sigma(\varphi)$. And combining this result with Postulate 5, we can obtain $f(x, \mathbf{A}) \supset_{fx} \Sigma(f(x))$.

Now, by the method of Case 3, we can prove $\Sigma \mathbf{x} \cdot \{\mathbf{F}\}(\mathbf{A})$ which is convertible into $\Sigma \mathbf{x} . \{\lambda \mathbf{x} . \mathbf{F}\}(\mathbf{x}, \mathbf{A})$. Combining this with the preceding, we can obtain $\{\lambda \mathbf{x} . \mathbf{F}\}(\mathbf{x}, \mathbf{A}) \supset_{\mathbf{x}} \Sigma(\{\lambda \mathbf{x} . \mathbf{F}\}(\mathbf{x}))$. And this is convertible into $\{\mathbf{F}\}(\mathbf{A}) \supset_{\mathbf{x}} \Sigma(\mathbf{F})$.

Thus we are able to prove $\Pi(\mathbf{F}', \mathbf{G}')$ where $\mathbf{F}'$ stands for $\lambda \mathbf{x} . \{\mathbf{F}\}(\mathbf{A})$ and $\mathbf{G}'$ stands for $\lambda \mathbf{x} \Sigma(\mathbf{F})$. And, by the hypothesis of our induction, we are able to prove $[\mathbf{M}] \supset_{\mathbf{x}} \{\mathbf{F}'\}(\mathbf{x})$. Therefore, by the method of Case 1, we can prove $[\mathbf{M}] \supset_{\mathbf{x}} \{\mathbf{G}'\}(\mathbf{x})$, and from it by conversion, $[\mathbf{M}] \supset_{\mathbf{x}} . \mathbf{N}$.

Case 6, the last step in the proof of **N** as a consequence of **M** is an application of Rule IV, and **F** contains **x** as a free variable, and **A** is identical with **x**. By hypothesis, $\Sigma \mathbf{x} . \mathbf{M}$ is provable, and, by the hypothesis of our induction, $[\mathbf{M}] \supset_{\mathbf{x}} \{\mathbf{F}\}(\mathbf{x})$ is provable. Therefore, by the method of Case 3, we can prove $\Sigma \mathbf{x} . \{\mathbf{F}\}(\mathbf{x})$, and this is convertible into $\Sigma \mathbf{x} . \{\lambda \mathbf{x} . \mathbf{F}\}(\mathbf{x}, \mathbf{x})$. Hence, using Postulate 11, we can obtain $\{\mathbf{F}\}(\mathbf{x}) \supset_{\mathbf{x}} \Sigma(\mathbf{F})$.

Thus we are able to prove $\Pi(\mathbf{F}', \mathbf{G}')$ where $\mathbf{F}'$ stands for $\lambda \mathbf{x} . \{\mathbf{F}\}(\mathbf{x})$ and $\mathbf{G}'$ stands for $\lambda \mathbf{x} \Sigma(\mathbf{F})$. And, by the hypothesis of our induction, we are able to prove $[\mathbf{M}] \supset_{\mathbf{x}} \{\mathbf{F}'\}(\mathbf{x})$. Therefore, by the method of Case 1, we can prove $[\mathbf{M}] \supset_{\mathbf{x}} \{\mathbf{G}'\}(\mathbf{x})$, and from it by conversion, $[\mathbf{M}] \supset_{\mathbf{x}} . \mathbf{N}$.

Case 7, the last step in the proof of **N** as a consequence of **M** is an application of Rule IV, and both **F** and **A** contain **x** as a free variable. By the hypothesis of our induction, $[\mathbf{M}] \supset_{\mathbf{x}} \{\mathbf{F}\}(\mathbf{A})$ is provable, and this is convertible into $[\mathbf{M}] \supset_{\mathbf{x}} \{\lambda y \{\mathbf{F}\}(\{\lambda \mathbf{x} . \mathbf{A}\}(y))\}(\mathbf{x})$. Therefore, by the method of Case 6, we can prove $[\mathbf{M}] \supset_{\mathbf{x}} \Sigma y . \{\mathbf{F}\}(\{\lambda \mathbf{x} . \mathbf{A}\}(y))$.

This last formula is convertible into $[\mathbf{M}] \supset_{\mathbf{x}} \{\lambda \varphi \Sigma y . \varphi(\{\lambda \mathbf{x} . \mathbf{A}\}(y))\}(\mathbf{F})$. Therefore, by the method of Case 4, we can prove $\Sigma \varphi \Sigma y . \varphi(\{\lambda \mathbf{x} . \mathbf{A}\}(y))$. And, combining this with Postulate 10, we can obtain $\Sigma y [\varphi(\{\lambda \mathbf{x} . \mathbf{A}\}(y))] \supset_{\varphi} \Sigma(\varphi)$. And hence, combining with Postulate 5, we can obtain

$$\Sigma y . [f(x, \{\lambda \mathbf{x} . \mathbf{A}\}(y))] \supset_{fx} \Sigma(f(x)).$$

By the method of Case 3, using $[\mathbf{M}] \supset_{\mathbf{x}} \Sigma y . \{\mathbf{F}\}(\{\lambda \mathbf{x} . \mathbf{A}\}(y))$, we can prove $\Sigma \mathbf{x} \Sigma y . \{\mathbf{F}\}(\{\lambda \mathbf{x} . \mathbf{A}\}(y))$ and this is convertible into $\Sigma x \Sigma y . \{\lambda \mathbf{x} . \mathbf{F}\}(x, \{\lambda \mathbf{x} . \mathbf{A}\}(y))$. Combining this last formula with the formula at the end of the preceding

paragraph, we can obtain $\Sigma y[\{\lambda \mathbf{x} \cdot \mathbf{F}\}(x, \{\lambda \mathbf{x} \cdot \mathbf{A}\}(y))] \supset_x \Sigma(\{\lambda \mathbf{x} \cdot \mathbf{F}\}(x))$. And this is convertible into $\Sigma y[\{\mathbf{F}\}(\{\lambda \mathbf{x} \cdot \mathbf{A}\}(y))] \supset_x \Sigma(\mathbf{F})$.

Thus we are able to prove $\Pi(\mathbf{F}', \mathbf{G}')$ and $[\mathbf{M}] \supset_x \{\mathbf{F}'\}(\mathbf{x})$ where $\mathbf{F}'$ stands for $\lambda \mathbf{x} \Sigma y \cdot \{\mathbf{F}\}(\{\lambda \mathbf{x} \cdot \mathbf{A}\}(y))$ and $\mathbf{G}'$ stands for $\lambda \mathbf{x} \Sigma(\mathbf{F})$. Therefore, by the method of Case 1, we can prove $[\mathbf{M}] \supset_x \{\mathbf{G}'\}(\mathbf{x})$, and from it by conversion, $[\mathbf{M}] \supset_x \cdot \mathbf{N}$.

Case 8, the last step in the proof of $\mathbf{N}$ as a consequence of $\mathbf{M}$ is an application of Rule V, and $\mathbf{A}$ contains $\mathbf{x}$ as a free variable, and $\mathbf{F}$ and $\mathbf{G}$ do not. By the hypothesis of our induction, $[\mathbf{M}] \supset_x \{\mathbf{F}\}(\mathbf{A})$ is provable. Hence, by the method of Case 4, we can prove $\Sigma(\mathbf{F})$. And, by the method of Case 3, we can prove $\Sigma \mathbf{x} \cdot \{\mathbf{F}\}(\mathbf{A})$, which is convertible into $\Sigma \mathbf{x} \cdot \{\mathbf{F}\}(\{\lambda \mathbf{x} \cdot \mathbf{A}\}(\mathbf{x}))$.

Also, by the hypothesis of our induction, $\Pi(\mathbf{F}, \mathbf{G})$ is provable. Hence, using Postulate 5, we can obtain $\{\mathbf{F}\}(f(x)) \supset_{fx} \{\mathbf{G}\}(f(x))$ and from this, using $\Sigma \mathbf{x} \cdot \{\mathbf{F}\}(\{\lambda \mathbf{x} \cdot \mathbf{A}\}(\mathbf{x}))$, we can obtain $\{\mathbf{F}\}(\{\lambda \mathbf{x} \cdot \mathbf{A}\}(\mathbf{x})) \supset_x \{\mathbf{G}\}(\{\lambda \mathbf{x} \cdot \mathbf{A}\}(\mathbf{x}))$.

From this last formula, using Rule II, we can obtain $\Pi(\mathbf{F}', \mathbf{G}')$, where $\mathbf{F}'$ and $\mathbf{G}'$ stand respectively for $\lambda \mathbf{x} \cdot \{\mathbf{F}\}(\mathbf{A})$ and $\lambda \mathbf{x} \cdot \{\mathbf{G}\}(\mathbf{A})$. And, applying Rule III to $[\mathbf{M}] \supset_x \{\mathbf{F}\}(\mathbf{A})$, we can prove $[\mathbf{M}] \supset_x \{\mathbf{F}'\}(\mathbf{x})$. Therefore, by the method of Case 1, we can prove $[\mathbf{M}] \supset_x \{\mathbf{G}'\}(\mathbf{x})$ and from it by conversion, $[\mathbf{M}] \supset_x \cdot \mathbf{N}$.

Case 9, the last step in the proof of $\mathbf{N}$ as a consequence of $\mathbf{M}$ is an application of Rule V, and $\mathbf{G}$ contains $\mathbf{x}$ as a free variable, and $\mathbf{F}$ and $\mathbf{A}$ do not. By the hypothesis of our induction, $\{\mathbf{F}\}(\mathbf{A})$ is provable, and hence, using Postulate 2, $\Pi(\mathbf{F}, \psi) \supset_\psi \psi(\mathbf{A})$ is provable.

Also, by the hypothesis of our induction, $[\mathbf{M}] \supset_x \Pi(\mathbf{F}, \mathbf{G})$ is provable, and from this by conversion we can obtain $[\mathbf{M}] \supset_x \{\lambda \psi \Pi(\mathbf{F}, \psi)\}(\mathbf{G})$.

Thus we can prove $\Pi(\mathbf{F}', \mathbf{G}')$ and $[\mathbf{M}] \supset_x \{\mathbf{F}'\}(\mathbf{A}')$ where $\mathbf{F}'$ stands for $\lambda \psi \Pi(\mathbf{F}, \psi)$ and $\mathbf{G}'$ stands for $\lambda \psi \cdot \psi(\mathbf{A})$, and $\mathbf{A}'$ stands for $\mathbf{G}$. Therefore, by the method of Case 8, we can prove $[\mathbf{M}] \supset_x \{\mathbf{G}'\}(\mathbf{A}')$ and from it by conversion, $[\mathbf{M}] \supset_x \cdot \mathbf{N}$.

Case 10, the last step in the proof of $\mathbf{N}$ as a consequence of $\mathbf{M}$ is an application of Rule V, and $\mathbf{F}$ and $\mathbf{G}$ contain $\mathbf{x}$ as a free variable, and $\mathbf{A}$ does not. We can prove $[\mathbf{M}] \supset_x \cdot \mathbf{N}$ as follows:

$\mathbf{B}_1$: $\Sigma \mathbf{x} \cdot \mathbf{M}$ —provable, by hypothesis.

$\mathbf{A}_1$: $\{\lambda \varrho \Sigma(\varrho)\}(\lambda \mathbf{x} \cdot \mathbf{M})$ —III, $\mathbf{B}_1$.

$\mathbf{A}_2$: $\Sigma(\varrho) \supset_\varrho \cdot \Sigma y[\varrho(x) \supset_x \varphi(x, y)] \supset_\varphi \cdot [\varrho(x) \supset_x \Pi(\varphi(x), \psi(x))] \supset_\psi \cdot$
$[\varrho(x) \supset_x \varphi(x, y)] \supset_y \cdot \varrho(x) \supset_x \psi(x, y)$ —4.

$\mathbf{A}_3$: $\{\lambda \varrho \cdot \Sigma y[\varrho(x) \supset_x \varphi(x, y)] \supset_\varphi \cdot [\varrho(x) \supset_x \Pi(\varphi(x), \psi(x))] \supset_\psi \cdot$
$[\varrho(x) \supset_x \varphi(x, y)] \supset_y \cdot \varrho(x) \supset_x \psi(x, y)\}(\lambda \mathbf{x} \cdot \mathbf{M})$ —V, $\mathbf{A}_2$, $\mathbf{A}_1$.

$\mathbf{A}_4$: $\{\lambda\varrho \,.\, \Sigma y[\varrho(x) \supset_x \varphi(x, y)] \supset_\varphi \,.\, [\varrho(x) \supset_x \Pi(\varphi(x), \psi(x))] \supset_\psi \,.$
$[\varrho(x) \supset_x \varphi(x, y)] \supset_y \,.\, \varrho(x) \supset_x \psi(x, y)\}\,(\lambda x \,.\, M)$ —I, $\mathbf{A}_3$.

$\mathbf{A}_5$: $\Sigma y[\{\lambda x \,.\, M\}\,(x) \supset_x \varphi(x, y)] \supset_\varphi \,.\, [\{\lambda x \cdot M\}\,(x) \supset_x \Pi(\varphi(x), \psi(x))] \supset_\psi \,.$
$[\{\lambda x \,.\, M\}\,(x) \supset_x \varphi(x, y)] \supset_y \,.\, \{\lambda x \,.\, M\}\,(x) \supset_x \psi(x, y)$ —II, $\mathbf{A}_4$.

$\mathbf{A}_6$: $\Sigma y[[M] \supset_x \varphi(x, y)] \supset_\varphi \,.\, [\{\lambda x \,.\, M\}\,(x) \supset_x \Pi(\varphi(x), \psi(x)] \supset_\psi \,.$
$[\{\lambda x \,.\, M\}\,(x) \supset_x \varphi(x, y)] \supset_y \,.\, \{\lambda x \,.\, M\}\,(x) \supset_x \psi(x, y)$ —II, $\mathbf{A}_5$.

$\mathbf{A}_7$: $\Sigma y[[M] \supset_x \varphi(x, y)] \supset_\varphi \,.\, [[M] \supset_x \Pi(\varphi(x), \psi(x))] \supset_\psi \,.$
$[\{\lambda x \,.\, M\}\,(x) \supset_x \varphi(x, y)] \supset_y \,.\, \{\lambda x \,.\, M\}\,(x) \supset_x \psi(x, y)$ —II, $\mathbf{A}_6$.

$\mathbf{A}_8$: $\Sigma y[[M] \supset_x \varphi(x, y)] \supset_\varphi \,.\, [[M] \supset_x \Pi(\varphi(x), \psi(x))] \supset_\psi \,.$
$[[M] \supset_x \varphi(x, y)] \supset_y \,.\, \{\lambda x \,.\, M\}\,(x) \supset_x \psi(x, y)$ —II, $\mathbf{A}_7$.

$\mathbf{A}_9$: $\Sigma y[[M] \supset_x \varphi(x, y)] \supset_\varphi \,.\, [[M] \supset_x \Pi(\varphi(x), \psi(x))] \supset_\psi \,.$
$[[M] \supset_x \varphi(x, y)] \supset_y \,.\, [M] \supset_x \psi(x, y)$ —II, $\mathbf{A}_8$.

$\mathbf{B}_2$: $[M] \supset_x \{F\}\,(A)$ —provable by hypothesis of induction.

$\mathbf{A}_{10}$: $[M] \supset_x \{\lambda x \,.\, F\}\,(x, A)$ —III, $\mathbf{B}_2$.

$\mathbf{A}_{11}$: $\{\lambda y \,.\, [M] \supset_x \{\lambda x \,.\, F\}\,(x, y)\}\,(A)$ —III, $\mathbf{A}_{10}$.

$\mathbf{A}_{12}$: $\Sigma y \,.\, [M] \supset_x \{\lambda x \,.\, F\}\,(x, y)$ —IV, $\mathbf{A}_{11}$.

$\mathbf{A}_{13}$: $\{\lambda\varphi\, \Sigma y \,.\, [M] \supset_x \varphi(x, y)\}\,(\lambda x \,.\, F)$ —III, $\mathbf{A}_{12}$.

$\mathbf{A}_{14}$: $\{\lambda\varphi \,.\, [[M] \supset_x \Pi(\varphi(x), \psi(x))] \supset_\psi \,.\, [[M] \supset_x \varphi(x, y)] \supset_y \,.$
$[M] \supset_x \psi(x, y)\}\,(\lambda x \,.\, F)$ —V, $\mathbf{A}_9$, $\mathbf{A}_{13}$.

$\mathbf{A}_{15}$: $[[M] \supset_x \Pi(\{\lambda x \,.\, F\}\,(x), \psi(x))] \supset_\psi \,.\, [[M] \supset_x \{\lambda x \,.\, F\}\,(x, y)] \supset_y \,.$
$[M] \supset_x \psi(x, y)$ —II, $\mathbf{A}_{14}$.

$\mathbf{A}_{16}$: $[[M] \supset_x \Pi(F, \psi(x))] \supset_\psi \,.\, [[M] \supset_x \{\lambda x \,.\, F\}\,(x, y)] \supset_y \,.$
$[M] \supset_x \psi(x, y)$ —II, $\mathbf{A}_{15}$.

$\mathbf{A}_{17}$: $[[M] \supset_x \Pi(F, \psi(x))] \supset_\psi \,.\, [[M] \supset_x \{F\}\,(y)] \supset_y \,.\, [M] \supset_x \psi(x, y)$ —II, $\mathbf{A}_{16}$.

$\mathbf{B}_3$: $[M] \supset_x \Pi(F, G)$ —provable, by hypothesis of induction.

$\mathbf{A}_{18}$: $[M] \supset_x \Pi(F, \{\lambda x \,.\, G\}\,(x))$ —III, $\mathbf{B}_3$.

$\mathbf{A}_{19}$: $\{\lambda\psi \,.\, [M] \supset_x \Pi(F, \psi(x))\}\,(\lambda x \,.\, G)$ —III, $\mathbf{A}_{18}$.

$\mathbf{A}_{20}$: $\{\lambda\psi \,.\, [[M] \supset_x \{F\}\,(y)] \supset_y \,.\, [M] \supset_x \psi(x, y)\}\,(\lambda x \,.\, G)$ —V, $\mathbf{A}_{17}$, $\mathbf{A}_{19}$.

$\mathbf{A}_{21}$: $[[M] \supset_x \{F\}\,(y)] \supset_y \,.\, [M] \supset_x \{\lambda x \,.\, G\}\,(x, y)$ —II, $\mathbf{A}_{20}$.

$\mathbf{A}_{22}$: $[[M] \supset_x \{F\}\,(y)] \supset_y \,.\, [M] \supset_x \{G\}\,(y)$ —II, $\mathbf{A}_{21}$.

$\mathbf{A}_{23}$: $\{\lambda y \,.\, [M] \supset_x \{F\}\,(y)\}\,(A)$ —III, $\mathbf{B}_2$.

$\mathbf{A}_{24}$: $\{\lambda y \,.\, [M] \supset_x \{G\}\,(y)\}\,(A)$ —V, $\mathbf{A}_{22}$, $\mathbf{A}_{23}$.

$\mathbf{A}_{25}$: $[M] \supset_x \{G\}\,(A)$ —II, $\mathbf{A}_{24}$.

$\mathbf{A}_{26}$: $[M] \supset_x \,.\, N$ —by conversion, from $\mathbf{A}_{25}$.

Case 11, the last step in the proof of **N** as a consequence of **M** is an application of Rule V, and **F** and **G** and **A** all three contain **x** as a free

variable. We include, in particular, the case that **F** and **G** both contain **x** as a free variable and **A** is identical with **x**, since there is nothing to be gained by treating this case separately. By the hypothesis of our induction, $[\mathbf{M}] \supset_{\mathbf{x}} \{\mathbf{F}\} (\mathbf{A})$ is provable, and this is convertible into

$$[\mathbf{M}] \supset_{\mathbf{x}} \{\lambda x \{\lambda y . \mathrm{S}^{\mathbf{x}}_{y} \mathbf{F}|\} (x, \{\lambda y . \mathrm{S}^{\mathbf{x}}_{y} \mathbf{A}|\} (x))\} (\mathbf{x}).$$

Therefore, by the method of Case 3, we can prove $\Sigma x \{\lambda y . \mathrm{S}^{\mathbf{x}}_{y} \mathbf{F}|\} (x, \{\lambda y . \mathrm{S}^{\mathbf{x}}_{y} \mathbf{A}|\} (x))$. Hence, using Postulate 7, we can prove

$$\{\lambda y . \mathrm{S}^{\mathbf{x}}_{y} \mathbf{F}|\} (x, \{\lambda y . \mathrm{S}^{\mathbf{x}}_{y} \mathbf{A}|\} (x)) \supset_{x} . \Pi (\{\lambda y . \mathrm{S}^{\mathbf{x}}_{y} \mathbf{F}|\} (x), \psi(x)) \supset_{\psi} \psi (x, \{\lambda y . \mathrm{S}^{\mathbf{x}}_{y} \mathbf{A}|\} (x)),$$

and this is convertible into

$$\{\mathbf{F}\} (\mathbf{A}) \supset_{\mathbf{x}} . \Pi (\mathbf{F}, \psi(\mathbf{x})) \supset_{\psi} \psi (\mathbf{x}, \mathbf{A}).$$

And hence, by the method of Case 1, since $[\mathbf{M}] \supset_{\mathbf{x}} \{\mathbf{F}\} (\mathbf{A})$ is provable, we can prove $[\mathbf{M}] \supset_{\mathbf{x}} . \Pi (\mathbf{F}, \psi(\mathbf{x})) \supset_{\psi} \psi (\mathbf{x}, \mathbf{A})$.

Thus we are able to prove $[\mathbf{M}] \supset_{\mathbf{x}} \Pi (\mathbf{F}', \mathbf{G}')$ where $\mathbf{F}'$ stands for $\lambda \psi \Pi (\mathbf{F}, \psi(\mathbf{x}))$ and $\mathbf{G}'$ stands for $\lambda \psi . \psi (\mathbf{x}, \mathbf{A})$. And, by the hypothesis of our induction, we are able to prove $[\mathbf{M}] \supset_{\mathbf{x}} \{\mathbf{F}'\} (\mathbf{A}')$, where $\mathbf{A}'$ stands for $\lambda x . \mathbf{G}$. Therefore, by the method of Case 10, we can prove $[\mathbf{M}] \supset_{\mathbf{x}} \{\mathbf{G}'\} (\mathbf{A}')$ and from it by conversion, $[\mathbf{M}] \supset_{\mathbf{x}} . \mathbf{N}$.

Case 12, the last step in the proof of **N** as a consequence of **M** is an application of Rule V, and **F** and **A** both contain **x** as a free variable, and **G** does not contain **x** as a free variable. As before, **A** may, in particular, be identical with **x**. By the hypothesis of our induction, $[\mathbf{M}] \supset_{\mathbf{x}} \{\mathbf{F}\} (\mathbf{A})$ is provable, and this is convertible into $[\mathbf{M}] \supset_{\mathbf{x}} \{\lambda \varphi . \varphi (\mathbf{A})\} (\mathbf{F})$. Therefore, by the method of Case 7, we can prove $[\mathbf{M}] \supset_{\mathbf{x}} \Sigma \varphi . \varphi (\mathbf{A})$, and this is convertible into $[\mathbf{M}] \supset_{\mathbf{x}} \{\lambda x E(x)\} (\mathbf{A})$. Hence, using Postulate 2, and the method of Case 8, we can prove $[\mathbf{M}] \supset_{\mathbf{x}} . \varphi (\mathbf{A}) \supset_{\varphi} . \Pi (\varphi, \psi) \supset_{\psi} \psi (\mathbf{A})$. And hence, by the method of Case 11, since $[\mathbf{M}] \supset_{\mathbf{x}} \{\mathbf{F}\} (\mathbf{A})$ is provable, we can prove $[\mathbf{M}] \supset_{\mathbf{x}} . \Pi (\mathbf{F}, \psi) \supset_{\psi} \psi (\mathbf{A})$.

Thus we are able to prove $[\mathbf{M}] \supset_{\mathbf{x}} \Pi (\mathbf{F}', \mathbf{G}')$ where $\mathbf{F}'$ stands for $\lambda \psi \Pi (\mathbf{F}, \psi)$ and $\mathbf{G}'$ stands for $\lambda \psi . \psi (\mathbf{A})$. And, by the hypothesis of our induction, we are able to prove $[\mathbf{M}] \supset_{\mathbf{x}} \{\mathbf{F}'\} (\mathbf{A}')$ where $\mathbf{A}'$ stands for $\mathbf{G}$. Therefore, by the method of Case 10, we can prove $[\mathbf{M}] \supset_{\mathbf{x}} \{\mathbf{G}'\} (\mathbf{A}')$, and from it by conversion, $[\mathbf{M}] \supset_{\mathbf{x}} . \mathbf{N}$.

Case 13, the last step in the proof of **N** as a consequence of **M** is an application of Rule V, and **A** is identical with **x**, **G** contains **x** as a free variable and **F** does not. By the hypothesis of our induction, $[\mathbf{M}] \supset_{\mathbf{x}} \{\mathbf{F}\} (\mathbf{x})$ is provable, and this is convertible into $[\mathbf{M}] \supset_{\mathbf{x}} \{\lambda \varphi . \varphi (\mathbf{x})\} (\mathbf{F})$. Therefore, by the method of Case 5, we can prove $[\mathbf{M}] \supset_{\mathbf{x}} \Sigma \varphi . \varphi (\mathbf{x})$ and this is con-

vertible into $[\mathbf{M}]\supset_{\mathbf{x}} E(\mathbf{x})$. Hence, using Postulate 6, and the method of Case 1, we can prove $[\mathbf{M}]\supset_{\mathbf{x}} . \varphi(\mathbf{x})\supset_{\varphi} . \Pi(\varphi, \psi(\mathbf{x}))\supset_{\psi} \psi(\mathbf{x}, \mathbf{x})$. And hence, by the method of Case 10, since $[\mathbf{M}]\supset_{\mathbf{x}} \{\mathbf{F}\}(\mathbf{x})$ is provable, we can prove $[\mathbf{M}]\supset_{\mathbf{x}} . \Pi(\mathbf{F}, \psi(\mathbf{x}))\supset_{\psi} \psi(\mathbf{x}, \mathbf{x})$.

Thus we are able to prove $[\mathbf{M}]\supset_{\mathbf{x}} \Pi(\mathbf{F}', \mathbf{G}')$ where $\mathbf{F}'$ stands for $\lambda\psi\,\Pi(\mathbf{F}, \psi(\mathbf{x}))$, and $\mathbf{G}'$ stands for $\lambda\psi . \psi(\mathbf{x}, \mathbf{x})$. And, by the hypothesis of our induction, we are able to prove $[\mathbf{M}]\supset_{\mathbf{x}} \{\mathbf{F}'\}(\mathbf{A}')$ where $\mathbf{A}'$ stands for $\lambda\mathbf{x} . \mathbf{G}$. Therefore, by the method of Case 10, we can prove $[\mathbf{M}]\supset_{\mathbf{x}} \{\mathbf{G}'\}(\mathbf{A}')$ and from it by conversion, $[\mathbf{M}]\supset_{\mathbf{x}} . \mathbf{N}$.

Case 14, the last step in the proof of **N** as a consequence of **M** is an application of Rule V, and **G** and **A** both contain **x** as a free variable, and **F** does not contain **x** as a free variable. By the hypothesis of our induction, we can prove $[\mathbf{M}]\supset_{\mathbf{x}} \{\mathbf{F}\}(\mathbf{A})$ and hence, by the method of Case 4, we can prove $\Sigma(\mathbf{F})$. Hence, using Postulate 5, we can prove $\Pi(\mathbf{F}, \psi)\supset_{\psi} . \{\mathbf{F}\}(f(x))\supset_{fx} \psi(f(x))$. And, by the hypothesis of our induction, we can prove $[\mathbf{M}]\supset_{\mathbf{x}} \Pi(\mathbf{F}, \mathbf{G})$, which is convertible into $[\mathbf{M}]\supset_{\mathbf{x}} \{\lambda_{\psi}\Pi(\mathbf{F}, \psi)\}(\mathbf{G})$. Therefore, by the method of Case 8, we can prove

$$[\mathbf{M}]\supset_{\mathbf{x}} \{\lambda_{\psi} . \{\mathbf{F}\}(f(x))\supset_{fx} \psi(f(x))\}(\mathbf{G}),$$

which is convertible into $[\mathbf{M}]\supset_{\mathbf{x}} . \{\mathbf{F}\}(f(x))\supset_{fx} \{\mathbf{G}\}(f(x))$.

Moreover, by conversion from $[\mathbf{M}]\supset_{\mathbf{x}} \{\mathbf{F}\}(\mathbf{A})$ we can prove

$$[\mathbf{M}]\supset_{\mathbf{x}} \{\lambda x . \{\mathbf{F}\}(\{\lambda\mathbf{x} . \mathbf{A}\}(x))\}(\mathbf{x}),$$

and therefore, by the method of Case 3, we can prove $\Sigma x . \{\mathbf{F}\}(\{\lambda\mathbf{x} . \mathbf{A}\}(x))$. Hence, by the method of Case 9, using the formula at the end of the preceding paragraph, we can prove

$$[\mathbf{M}]\supset_{\mathbf{x}} . \{\mathbf{F}\}(\{\lambda\mathbf{x} . \mathbf{A}\}(x))\supset_{x} \{\mathbf{G}\}(\{\lambda\mathbf{x} . \mathbf{A}\}(x)).$$

Hence, using $[\mathbf{M}]\supset_{\mathbf{x}} \{\lambda x . \{\mathbf{F}\}(\{\lambda\mathbf{x} . \mathbf{A}\}(x))\}(\mathbf{x})$ and the method of Case 13, we can prove

$$[\mathbf{M}]\supset_{\mathbf{x}} \{\lambda x . \{\mathbf{G}\}(\{\lambda\mathbf{x} . \mathbf{A}\}(x))\}(\mathbf{x}).$$

And hence finally, by conversion, we can prove $[\mathbf{M}]\supset_{\mathbf{x}} . \mathbf{N}$.

Case 15, the last step in the proof of **N** as a consequence of **M** is an application of Rule V, and neither **F** nor **G** nor **A** contains **x** as a free variable. By the hypothesis of our induction, $\Pi(\mathbf{F}, \mathbf{G})$ and $\{\mathbf{F}\}(\mathbf{A})$ are both provable. Hence, using Rule V, $\{\mathbf{G}\}(\mathbf{A})$ is provable, and from it by conversion, **N**.

Case 16, the last step in the proof of **N** as a consequence of **M** is an application of Rule IV, and neither **F** nor **A** contains **x** as a free variable.

By the hypothesis of our induction, $\{\mathbf{F}\}(\mathbf{A})$ is provable. Hence, using Rule IV, $\Sigma(\mathbf{F})$ is provable, and from it by conversion, **N**.

COROLLARY. *Suppose that* **M** *contains* **x** *and* **y** *as free variables and that* $\Sigma\mathbf{x}\,\Sigma\mathbf{y}\,.\,\mathbf{M}$ *is provable, and that* **N** *is provable as a consequence of* **M**. *Then if* **N** *contains neither* **x** *nor* **y** *as a free variable, the formula* **N** *is a provable formula. If* **N** *contains* **y** *as a free variable, the formula* $[\mathbf{M}]\supset_{\mathbf{xy}}.\,\mathbf{N}$ *is a provable formula. And if* **N** *contains* **x** *as a free variable, the formula* $[\mathbf{M}]\supset_{\mathbf{yx}}.\,\mathbf{N}$ *is a provable formula.*

And similarly for cases where **M** cotains a greater number of free variables.

THEOREM II. *If* $\Sigma(\mathbf{F})$ *is provable, and* **G** *does not contain* **x** *as a free variable, and* $\{\mathbf{G}\}(\mathbf{x})$ *is provable as a consequence of* $\{\mathbf{F}\}(\mathbf{x})$, *then* $\Pi(\mathbf{F},\mathbf{G})$ *is provable.*

For, using Postulate 12, we can prove $\Sigma\mathbf{x}\,.\,\{\mathbf{F}\}(\mathbf{x})$. Therefore, by Theorem I, $\{\mathbf{F}\}(\mathbf{x})\supset_{\mathbf{x}}\{\mathbf{G}\}(\mathbf{x})$ is provable. And hence, using Postulate 13, we can prove $\Pi(\mathbf{F},\mathbf{G})$.

THEOREM III. *If* $\Sigma(\mathbf{F})$ *is provable, and* **N** *does not contain* **x** *as a free variable, and* **N** *is provable as a consequence of* $\{\mathbf{F}\}(\mathbf{x})$, *then* **N** *is provable.*

For, using Postulate 12, we can prove $\Sigma\mathbf{x}\,.\,\{\mathbf{F}\}(\mathbf{x})$. Therefore, by Theorem I, **N** is provable.

Moreover it is readily seen that if our rules of procedure are to be Rules I—V and if Theorems II and III are to be true of our system then Postulates 1—13 must be, if not postulates (as we have taken them to be), at least provable formulas.

PRINCETON UNIVERSITY, PRINCETON, N. J.

CIRCUITS UPON POLYHEDRA.*

By Clarence N. Reynolds, jr.

Introduction. In this paper we shall restrict ourselves to the consideration of figures homeomorphic with simply connected convex polyhedra. That is to say, we shall be concerned with maps of sets of regions on a simply connected closed surface subject to the conditions that (*a*) no region shall be multiply connected and (*b*) no pair of adjoining regions shall form a multiply connected whole.

With reference to these maps, we shall study the conditions under which it may be possible to introduce closed unicursal curves which either (*a*) pass once and only once through each region of our map or (*b*) pass along some of the boundaries of our map in such a way as to pass once and only once through every vertex of our map. These closed curves, when oriented, will define sequences of regions, or vertices, which we shall refer to as complete cyclic sequences of regions, or of vertices.

In a recent paper[1] by Hassler Whitney, he has shown that complete cyclic sequences of regions exist in all cubic polyhedral maps containing no multiply connected sets of three regions, and therefore no triangles. Our theorems are less significant since they are reduction theorems, reducing a question concerning a given map to a similar question concerning a map which may be almost as difficult to work with as is the original map. Referring, however, to Brückner's catalog of cubic convex polyhedra[2] we find that he lists $n(f)$ polyhedra of f faces each, of which $t(f)$ polyhedra contain one or more triangles, where $n(4) = t(4) = 1$; $n(5) = t(5) = 1$; $n(6) = 2$, $t(6) = 1$; $n(7) = 5$, $t(7) = 4$; $n(8) = 14$, $t(8) = 12$; $n(9) = 50$, $t(9) = 45$; $n(10) = 232$, $t(10) = 220$. Since our complete cyclic sequences of regions are found by experiment to exist in all of these 305 polyhedra and since 284 of these do not satisfy Dr. Whitney's hypothesis, we have remaining the task of studying what appears to be, judging from an *unfair* sample, a remainder of 93 per cent of the class of convex polyhedra.

After having derived some necessary and sufficient conditions for the existence of complete cyclic sequences, we shall apply our results to earlier studies of polyhedra, and to the problem of coloring in four colors the map of any set of regions on a simply connected closed surface.

1. **Cyclic sequences of regions.** When dealing with polyhedral maps of degree 3 (each of whose vertices is incident with three boundaries) we

* Received November 17, 1930, presented to the American Mathematical Society, September 9, 1930.

[1] Hassler Whitney, A theorem on graphs, Annals of Mathematics, vol. 32, p. 378.

[2] Vielecke und Vielflache, Teubner, Leipzig, 1900.

shall describe the immediate neighborhood of a boundary by indicating whether any one of its four neighboring regions is a region by which a given complete cyclic sequence of regions enters, leaves, enters and leaves, or does not enter nor leave the neighborhood of our boundary; by associating with our neighboring regions the symbols $+$, $-$, $\pm$, and 0 respectively. Thus, the neighborhood illustrated is described by the symbol $(+, -; 0, \pm)$ the symbols being associated with regions in the order $(a, b; c, d)$.

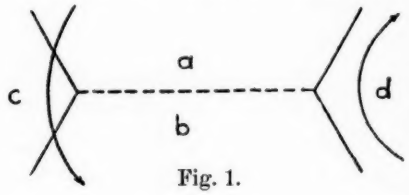

Fig. 1.

Two neighborhoods will, from our point of view be considered equivalent if the symbol of one can be obtained from that of the other (a) by interchanging the first two symbols, or the last two symbols; (b) by reversing the order of the four symbols or (c) by changing all of the signs (i. e. by changing the orientation of the directed sequence of regions).

Since with reference to any given complete cyclic sequence of regions, the neighborhood of any boundary must have as many points of egress as of ingress the number of plus signs in a given symbol must be equal to the number of minus signs in the symbol. Therefore, to secure a degree of symmetry in the matrix to be defined later, we shall order or alphabetize our symbols in the order $(+, \pm, 0, -)$ whenever dealing with the regions a and b and in the reverse order $(-, 0, \pm, +)$ when dealing with the regions c and d. We are now in a position to state

THEOREM 1. *The necessary and sufficient condition that a cubic polyhedral map of a regions shall contain a complete cyclic sequence of regions is that it contains at least one boundary, the erasure of which yields a cubic map of* $a-1$ *regions containing at least one complete cyclic sequence of regions with respect to which the deleted boundary is not of the type* $(\pm, 0; \pm, \pm)$.

Proof. The possible distinct types of neighborhoods are indicated in the accompanying matrix. Here impossible combinations of symbols are indicated by zeros when the impossibility is implied by unequal numbers of plus and minus signs in the symbol; and by crosses when it is found by experiment to be impossible to construct such a neighborhood for a complete sequence of regions in

(a, b \; c, d)	; −, −)	; −, 0)	; −, ±)	; −, +)	; 0, 0)	; 0, ±)	; ±, ±)
(+, +	a	0	0	0	0	0	0
(+, ±	0	×	a	0	0	0	0
(+, 0	0	a	$a^*(a-1)$	0	0	0	0
(+, −	0	0	0	×	×	a	$a^*(a-1)$
(±, ±	0	0	0	×	×	×	a
(±, 0	0	0	0	a	×	×	$(a-1)$
(0, 0	0	0	0	$a^*(a-1)$	a	×	×

either a complete map of a regions or in our deleted map of $a-1$ regions, without (a) missing some regions entirely (b) entering some regions with no possible exit, or (c) having two similarly directed arcs of a unicursal closed curve on a sphere, within each other's neighborhood.

Whenever a given combination of symbols is possible with reference to a complete cyclic sequence of regions in our original (reduced) map of a, $(a-1)$ regions, we have an a, $(a-1)$, entered in our matrix. Whenever a combination of symbols for a given boundary implies that the corresponding complete sequence crosses the boundary in question, that fact is indicated by a star in our matrix.

From our matrix, we now see that if a complete cyclic sequence of regions exists in a cubic map of a regions then the erasure of an edge crossed by our circuit yields a map of $a-1$ regions containing a complete cyclic sequence of regions with respect to which the deleted boundary is *not* of the type $(\pm, 0; \pm, \pm)$. Conversely, if the erasure of any boundary of a cubic map of a regions yields a map containing a complete cyclic sequence of regions with respect to which the neighborhood of our deleted boundary is not of the type $(\pm, 0; \pm, \pm)$ then the replacement of the boundary makes possible the construction of a complete cyclic sequence of regions in our original map.

2. **Cyclic sequences of vertices.** We now develop a somewhat trivial theorem similar to the preceding theorem and applicable to complete cyclic sequences of vertices. This theorem will be proven by a similar process of erasing or deleting boundaries.

By an extended neighborhood of the boundary we shall refer to the set of not more than thirteen boundaries each of which has at least one end point incident with the five lines which are in turn incident with the extremities of our deleted boundary. If a cubic map of a regions contains no boundary the erasure of which yields a cubic map of $a-1$ regions containing no complete cyclic sequence of vertices, then the complete cyclic sequence in the reduced map must enter and leave the extended neighborhood of our deleted boundary, either two, three, or four times. Let these types of extended neighborhoods be denoted by M_2, M_3 and M_4. We can now prove

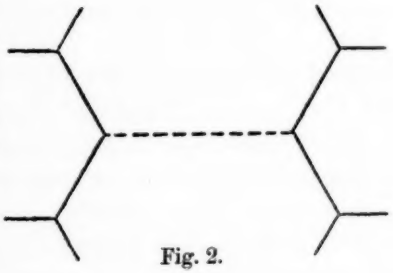

Fig. 2.

THEOREM 2. *A necessary and sufficient condition that a cubic map of a regions contain a complete cyclic sequence of vertices is that the map contain at least one boundary, the erasure of which yields a reduced map of* $(a-1)$

regions containing at least one complete cyclic sequence of vertices with respect to which the neighborhood of the deleted boundary is of type M_2.

Proof. To prove our condition necessary we have only to erase from a cubic map of a regions containing a complete cyclic sequence of vertices, a boundary which is not included between two consecutive vertices of our cyclic sequence of vertices. The resulting extended neighborhood will be of type M_2. The proof of sufficiency follows immediately from the definition since we can, in an extended neighborhood of type M_2 replace the erased boundary and the cyclic sequence of vertices which was complete with respect to the reduced map will be complete with respect to the original map.

3. Neighborhoods of regions. Turning now to the neighborhoods of regions, by which phrase we shall refer to the sets of distinct regions in contact with a single central region, we can describe the relation of such a neighborhood to any complete cyclic sequence of regions by means of the symbols $+$, $-$, $\pm$ and 0 just as we did in dealing with neighborhoods of boundaries. We can delete any of those polygons of fewer than six sides, of which some exist in any cubic map, by erasing one boundary. The possible types of resulting neighborhoods are the following:

$(0, 0, 0)$, $(+, -, 0)$, $(+, -, \pm)$, $(\pm, \pm, \pm)^*$; $(0, 0, 0, 0)$, $(+, -, 0, 0)$, $(+, 0, -, 0)$, $(+, -, +, -)$, $(+, -, \pm, 0)$, $(+, \pm, -, 0)$, $(+, -, \pm, \pm)$, $(+, \pm, -, \pm)$, $(\pm, \pm, \pm, \pm)^*$; $(0, 0, 0, 0, 0)$, $(+, -, 0, 0, 0)$, $(+, 0, -, 0, 0)$, $(+, -, +, -, 0)$, $(+, -, \pm, 0, 0)$, $(+, -, 0, \pm, 0)$, $(+, \pm, -, 0, 0)$, $(+, \pm, 0, -, 0)$, $(+, -, +, -, \pm)$, $(+, -, \pm, \pm, 0)$, $(+, -, \pm, 0, \pm)^*$, $(+, \pm, -, 0, \pm)$, $(+, \pm, \pm, -, 0)$, $(+, -, \pm, \pm, \pm)$, $(+, \pm, -, \pm, \pm)$, $(\pm, \pm, \pm, \pm, \pm)^*$.

Upon drawing all of the corresponding diagrams, we find it possible to find closed curves which define, simultaneously, cyclic sequences of regions in the original and in the reduced maps, in all cases excepting only those in the above list which are starred. Hence our

THEOREM 3. *The necessary and sufficient condition that a cubic map of a regions contain a complete cyclic sequence of regions is that there exist at least one boundary of a region of fewer than six sides, the erasure of which yields a cubic polyhedral map containing at least one complete cyclic sequence of regions with respect to which the neighborhood of the former "region of fewer than six sides" is not of one of the types*: $(\pm, \pm, \pm)$, $(\pm, \pm, \pm, \pm)$, $(\pm, \pm, \pm, \pm, \pm)$, *or* $(+, -, \pm, 0, \pm)$.

4. Irreducible maps. If we assume the completeness of the catalog of cubic polyhedra of no more than 10 faces published by Brückner in 1900[3] then we easily find experimentally that every such polyhedron

[3] Loc. cit. [2].

represents a map containing a complete cyclic sequence of regions. On the other hand, the eleven faced polyhedron (Fig. 3) obtained by truncating the six vertices of a triangular prism with triangular faces, contains no such cyclic sequence since the hypotheses of Theorems 1 and 3 are not satisfied by this figure which reduces to Brückner's polyhedron X_1.

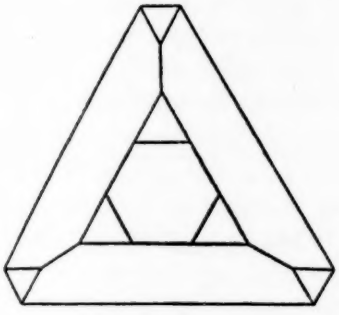

Fig. 3.

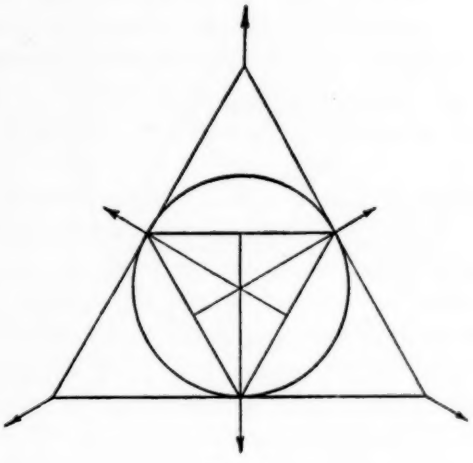

Fig. 3′.

The polar reciprocal of this map (Fig. 3′) is a map (non-cubic) containing no complete cyclic sequence of vertices. Every cubic vertex of this map is joined by each incident boundary to a non-cubic vertex. In this figure, and in Fig. 4′, arrowheads represent boundaries incident with a single vertex "on the other side of the sphere".

From our definitions it follows that if a given map contains no complete cyclic sequence of regions, its polar reciprocal map will be a map containing the same number of boundaries and containing no complete cyclic sequence of regions. Let us now consider the two types of sequences simultaneously and consider those maps with no complete sequence of regions (or vertices) and which have as few boundaries as any other such map. These maps, we shall, by analogy with the four-color problem, call irreducible maps with respect to complete cyclic sequence of regions (or vertices). In these maps we have no boundaries separating pairs of polygons each having more than three sides (joining pairs of vertices each of degree greater than three) for in such cases, the boundaries may be erased (contracted to a point) without producing polygons of two sides (vertices of degree two). Therefore, in these maps, any polygon of more than three sides (vertex of degree greater than three) is surrounded by triangles (vertices of degree three).

A second pair of polar reciprocal maps (represented in Figs. 4 and 4′) having no complete cyclic sequence of regions, or vertices, respectively, is due to Professor Orrin Frink. The map in Fig. 4 has fewer boundaries than that in Fig. 3, but, since its vertices are not all of degree 3, it does

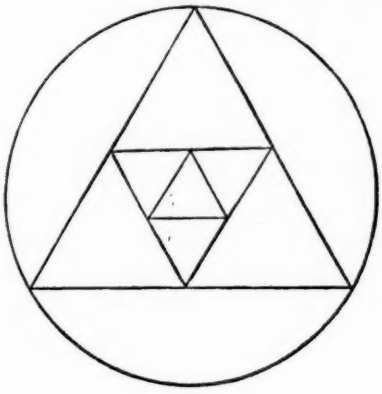

Fig. 4.

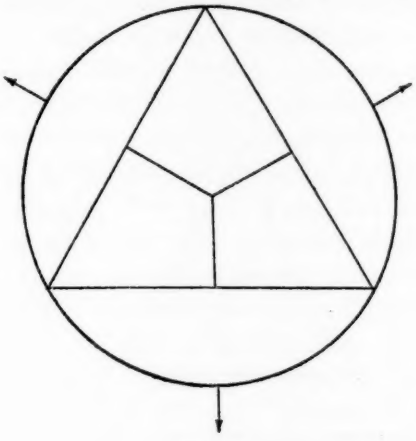

Fig. 4′.

not fall within the scope of our Theorems 1 and 3. Professor Frink's two maps do, however, possess the property just proven for irreducible maps. Nevertheless, they have not been proven irreducible. Assuming, as above, the accuracy of Brückner's catalog of cubic polyhedral maps, Fig. 3, is irreducible with respect to complete sequences of regions in cubic polyhedral maps.

5. **Application to the four-color problem.** The necessary and sufficient condition that it be possible to divide the regions of a cubic map into two simply connected sets of regions such that the regions of one set may be colored in two colors a and b while the regions of the other set are colored in two other colors c and d, is that the cubic map shall contain a complete cyclic sequence of vertices. This follows immediately from our definitions and seems to be well known. So far as known, every cubic map satisfies this condition. If there exist no cubic maps which are irreducible with respect to complete cyclic sequences of vertices, then any cubic map may be colored in four colors. It is known that if all cubic maps may be colored in four colors, then all maps of sets of regions on a simply connected closed surface may be so colored.[4] Thus we have another unsolved formulation of the four-color problem.

[4] For a list of the known reductions of the four-color problem, see the author's paper, "On the Problem of Coloring Maps in Four Colors, II", Annals of Mathematics, vol. 28, pp. 477–492.

WEST VIRGINIA UNIVERSITY, MORGANTOWN, W. VA.

GROUP INVARIANTS AND TORSION COEFFICIENTS.[1]

By Arthur B. Brown.

1. **Introduction.** The paper is largely expository. We consider two sets of invariants of finite Abelian groups; proofs of their uniqueness; relation to each other, to the Poincaré numbers of an infinite group with a finite number of generators, and to torsion coefficients in analysis situs. A possible alternative choice of the torsion coefficients is mentioned, not equivalent to the usual one.

2. **Invariants of finite Abelian groups.** Given a finite Abelian group G, a base[2] can be found whose elements E satisfy either (in general not both) of the following conditions: (a) the orders of the E's are powers of primes; (b) the order of any E divides the orders of all E's of higher orders. The orders of the E's will be called the *Gauss invariants* of G if (a) is satisfied; the *Schering invariants* if (b) is satisfied. The invariants so named were first described by Gauss[3] and Schering[4] respectively, but were first proved to be unique by Frobenius and Strickelberger.[5] The Schering invariants are called "the invariants" by Frobenius-Strickelberger, Bianchi, Hilton, Miller-Blichfeldt-Dickson; the Gauss invariants are called "the invariants" by Weber[6] and Speiser.[7]

The uniqueness of the Gauss and Schering invariants are established by similar proofs. We mention four brief proofs.

First proof. This proof makes use of a sub-group of elements which are fixed powers of elements of G. It was given by Weber,[8] and is found in the books of Weber, Bianchi, Hilton and Speiser.

---

[1] Received May 10 and November 1, 1931. Presented to the American Mathematical Society, September 8, 1931.

[2] The word "base" is used to denote a set of elements which generate G, and satisfy no relations other than those stating that they are of finite orders greater than unity. Cf. A. Speiser, "Die Theorie der Gruppen von endlicher Ordnung", second edition, Berlin, 1927, page 46.

[3] C. F. Gauss, "Démonstration de quelques théorèmes concernant les périodes des classes des formes binaires du second degré" (1801). Werke, vol. 2, pp. 266–268.

[4] E. Schering, "Die Fundamental-Classen der zusammensetzbaren arithmetischen Formen". Abh. d. Kgl. Ges. d. Wiss. zu Göttingen, vol. 14 (1869). Gesammelte Werke, pp. 135–148.

[5] G. Frobenius and L. Strickelberger, "Über Gruppen von vertauschbaren Elementen". Journ. f. d. reine u. angewandte Math. (Crelle Journal), vol. 86 (1879), pp. 217–262.

[6] H. Weber. "Algebra", vol. 2, second edition (1899), pp. 45–47.

[7] Loc. cit., p. 50.

[8] H. Weber, "Theorie der Abel'schen Zahlkörper". Acta Mathematica, vol. 9 (1887), pp. 105–130 (106–109).

The next three proofs were actually given in a more general case, mentioned in § 3, but also apply here.

Second proof. In this proof, starting with two different bases, we increase each step by step till it contains all the elements of the other. (The resulting "base" will of course not be a base in the sense described above.[2]) This proof was given by Tietze,[9] also by Alexander.[10]

Third proof. In this proof, due to Alexander,[11] one considers the number of elements which are independent "modulo m", for different values of m.

Fourth proof. This proof, depending on properties of linear independence, is in the second edition of Veblen's "Analysis Situs".[12]

REMARK 1. *If the order of a finite Abelian group is a power of a prime, the Gauss invariants are the same numbers as the Schering invariants.*

This follows from the definitions, and the fact that the order of any element of the group must be a power of the prime.

RULE 2. *Given the Schering invariants of a finite Abelian group, let each be factored into a product of powers of distinct primes. The totality of numbers so obtained are the Gauss invariants of the group.*

This result follows from the first part of Theorem 48 in Speiser, loc. cit.

For example, if the Schering invariants are 3, 3, 18, the Gauss invariants will be 2, 3, 3, 9.

RULE 3. *Given the Gauss invariants of a finite Abelian group, let D_r denote the highest common factor of all the products of r distinct (but not necessarily unequal) members of the set of invariants. Let $d_j = D_{j+1}/D_j$, $D_{j+1}, D_j \neq 0$. Those of the numbers d_j not equal to unity are then the Schering invariants of the group.*

Proof. This follows from the fact that the Schering invariants can be determined as the invariant factors of the matrix of powers (coefficients, under Veblen's notation) in a set of generating relations for an arbitrary set of elements which generate the group.[13]

REMARK 4. *While the Gauss invariants of the direct product of two finite Abelian groups are simply the set of numbers obtained by adjoining*

[9] H. Tietze, "Über die topologischen Invarianten mehrdimensionaler Mannigfaltigkeiten". Monatsh. f. Math. u. Phys., vol. 19 (1908), pp. 1–118 (62–63).

[10] J. W. Alexander, "Topological invariants of knots and links". Trans. Amer. Math. Soc., vol. 30 (1928), pp. 275–306 (293–295). Cf. S. Lefschetz, "Topology", New York (1931), pp. 31–32. (Lefschetz I.)

[11] J. W. Alexander, "Combinatorial analysis situs, I". Trans. Amer. Math. Soc., Vol. 28 (1926), pp. 301–329.

[12] O. Veblen, "Analysis Situs", second edition, pp. 128–129. New York, 1931. (Veblen I.)

[13] Tietze, loc. cit., page 56. Cf. Veblen I, pp. 150–151; Lefschetz I, pp. 29–30.

the Gauss invariants of the one factor group to those of the other, the corresponding result does not hold for the Schering invariants.

This follows easily from the definitions.

3. **Poincaré numbers and torsion coefficients.** Given a group G with a finite number of generators, let $\bar{G}$ be the Abelian group determined from G by making its elements permutable. Given any set of generators of $\bar{G}$ and a corresponding set of generating relations, the invariant factors of the matrix of exponents of the latter are called the *Poincaré numbers* of G.[14] From the form of the generating relations corresponding to these invariant factors, it follows that the Poincaré numbers of G are the Schering invariants of the sub-group H of $\bar{G}$ consisting of all its elements of finite orders. Consequently, as remarked by Veblen,[15] the uniqueness of the Poincaré numbers of G is a corollary of the uniqueness of the Schering invariants of H. (The second, third and fourth proofs mentioned in § 2, all deal directly with the Abelian group $\bar{G}$.)

The k-dimensional torsion coefficients of a complex in analysis situs are the Poincaré numbers of the (Abelian) k-th "homology group" of the complex.[16] As the homology groups are topological invariants of the complex, the invariance of the torsion coefficients is therefore a consequence of the uniqueness of the Poincaré numbers of the homology groups. From our remark above connecting the latter with Schering invariants, it follows that in § 2 we have mentioned the equivalent of four short proofs of the group-theoretic part of the proof of the invariance of the coefficients of torsion.

4. **A rule for determining the invariants.** It was shown by Schering[17] that if A is any element of highest possible order belonging to a finite Abelian group G, a base for G can be found, one of whose elements is A, and the orders of whose elements satisfy the definition of the Schering invariants of G. This fact gives rise to the following algorithm, suggested by Veblen.

RULE 5. *Given a finite Abelian group G_1, let A_1 be any element of highest possible order, H_1 the group generated by A_1, and G_2 the quotient-group G_1/H_1.*

[14] H. Tietze, loc. cit., pp. 56, 58, 62.

[15] O. Veblen, "Invariance of the Poincaré numbers of a discrete group". Bull. Amer. Math. Soc., Vol. 30 (1924), pp. 405–406.

[16] Cf. W. Mayer, "Über abstrakte Topologie". Monatshefte für Math. u. Phys., vol. 36 (1929), pp. 1–42 (p. 11). The kth homology group is the group G_k defined in Veblen I, p. 128. Cf. also pp. 150–151, and Lefschetz I, p. 34. This statement partially answers the question raised in the last paragraph on p. 151 in Veblen I.

[17] Werke, pp. 139–140. Cf. L. Kronecker, "Auseinandersetzung einiger Eigenschaften der Klassenanzahl idealer complexer Zahlen". Berliner Monatsberichte, 1870, pp. 881–889; Collected Works, vol. 1, pp. 274–278.

Let A_2 be any element of G_2 of highest possible order, and so on. ⋯ Finally a group G_{s+1} will be reached, containing one element. Then the orders $r_1, r_2, \cdots, r_s$ of the elements $A_1, A_2, \cdots, A_s$ are the Schering invariants of G_1.

Proof. The result follows from the fact stated at the beginning of the section and the uniqueness of the Schering invariants. The following theorem can be used here in place of the uniqueness of the Schering invariants.

THEOREM 6. *Let H_1, H_1', H_2, H_2' be finite commutative groups; G_1 the direct product of H_1 and H_1', and G_2 the direct product of H_2 and H_2'. If G_1 and H_1 are isomorphic to G_2 and H_2 respectively, then H_1' must be isomorphic to H_2'.*

Proof. The result is an immediate consequence of the uniqueness of the Gauss invariants, and the fact stated in the first part of Remark 4.

REMARK. It might be thought that Schering and Kronecker therefore gave proofs of the uniqueness of the Schering invariants. However, this is not the case, as Theorem 6 was proved by use of the uniqueness of the invariants; and the result is not obvious, as the following fact shows: If G is a finite commutative group, and H_1 and H_2 two isomorphic sub-groups, then the quotient-groups G/H_1 and G/H_2 are not necessarily isomorphic.[18]

5. **Choice of the torsion coefficients.** We have seen (§ 3) that if H is the sub-group of all elements of finite order of the kth homology group, the k-dimensional torsion coefficients are the Schering invariants of H. The Gauss invariants of H might equally well be introduced as torsion coefficients, thus in general increasing their number. Such a procedure would have (1) the advantage that if two complexes K_1 and K_2 have no points in common, the torsion coefficients of $K = K_1 + K_2$ would be the set obtained by adjoining the torsion coefficients of K_1 to those of K_2 (cf. Remark 4); and (2) the disadvantage that the torsion coefficients would no longer be the invariant factors of the matrices of incidence relations of the complex (but cf. Rule 2).

[18] Cf. W. Burnside, "Theory of groups of finite order", second edition, 1911, § 29. A simple example is the following. Let G be the Abelian group generated by the elements A and B, of orders 2 and 4 respectively. Let H_1 be the sub-group consisting of A and the identity, and H_2 the sub-group consisting of B^2 and the identity. Then the quotient-groups G/H_1 and G/H_2 are not isomorphic. On the other hand, see Theorem 6.

COLUMBIA UNIVERSITY.

ON NORMAL PRODUCTS OF ALGEBRAS.[1]

By Jakob Levitzki.

1. **Introduction.** The direct multiplication of two algebras furnishes a method for obtaining new algebras from given ones. Another special method of multiplication leads to the construction of Dickson algebras. The purpose of the present paper is to study the properties of a generalized method of multiplication which includes as special cases those stated above. The product $P = A \cdot \bar{A}$ of two algebras A and $\bar{A}$ over the field F, which are subsets of a given algebra (this restriction, however, can be replaced by a postulational construction of the product) is called *normal with respect to* $\bar{A}$, if a basis $a_1, \cdots, a_n$ of A over F can be chosen such that $P = \sum_\nu a_\nu \bar{A}$ is a direct module over A, and $a_\nu \bar{A} = \bar{A} a_\nu$ for $\nu = 1, \cdots, n$. Every element b of A such that $b\bar{A} = \bar{A}b$, is said to be *normal over* A, and the automorphism $\bar{a} \to \bar{a}'$ of $\bar{A}$, where $b\bar{a} = \bar{a}'b$, is said to be *induced by* b. The set of all the automorphisms of $\bar{A}$ induced by normal elements of A (in short: induced by A) is finite, and does not exceed the order of the algebra A. If $\psi_1, \psi_2, \cdots, \psi_t$ are the different automorphisms of $\bar{A}$ induced by A, then the algebra A can be uniquely represented by a direct sum $A = \sum_1^t M_\lambda$, where M_λ is the set of all the normal elements of A which induce the automorphism ψ_λ of A. The modules M_λ are called the *automorphism-modules of* A *over* A. The set K of all the elements of A which are commutative with every element of A is called the *kernel of* A (over $\bar{A}$) and, if $K \neq 0$, it is an algebra over F and is identical with an M_λ. It follows that the algebra K is potent if, and only if the algebra K is potent. We obtain a one-to-one correspondence between the *normal primitive right (left) ideals* of A (i. e. minimal potent ideals of A which possess a normal basis with respect to $\bar{A}$) and the *primitive right (left) ideals* of K. We prove that A can be represented as a direct sum $A = \sum_\lambda A_\lambda + N$, where N is a *normal module over* $\bar{A}$ (i. e. N possesses a basis over F which is normal over $\bar{A}$) and a subset of the radical of A, while the A_λ are *normal-primary algebras* over A with units; that is, every ideal of each A_λ which is normal over $\bar{A}$ is equal to A_λ or nilpotent. Each normal primary algebra contains a set of n^2 matric units c_{ik}; $i, k = 1, \cdots, n$ which are normal over $\bar{A}$, and

[1] Received September 28, 1931. This paper was written while the author was a Sterling Research Fellow at Yale; it was presented to the Society April 3, 1931.

each $c_{ii} A_\lambda c_{ii}$ is *normal primitive* over $\bar{A}$, that is, every normal right or left ideal of $c_{ii} A_\lambda c_{ii}$ is equal to $c_{ii} A_\lambda c_{ii}$ or nilpotent. We prove further that every normal product can be represented in the form $P = \sum_\lambda A_\lambda^* \bar{A}_\lambda + N\bar{A}$ where the $\bar{A}_\lambda$ are isomorphic to each other and to $\bar{A}$, and $A = \sum_\lambda A_\lambda^* + N$. Here is N a subset of the radical of A and a normal module over A, while each algebra A_λ^* is a normal primary algebra over $\bar{A}$ and over the algebra $\bar{A}_\lambda$. Further we show that $A_\alpha^* = C_\alpha \times A_\alpha^{**}$, where C_α is a total matric algebra over F and its elements are commutative with every element of $\bar{A}_\alpha$, and where each A_α^{**} is normal primitive over $\bar{A}_\alpha$ and its automorphism-modules form a group. Thus the further study of normal products is reduced to the case where A is normal primitive over $\bar{A}$ and its automorphism-modules form a group. If K denotes the kernel of A, then it follows that A is a *group algebra modulo the subalgebra* K, that is, the algebra A contains a system of regular elements $a_1, \cdots, a_m$ such that $A = \sum a_\lambda K$; $a_\lambda K = K a_\lambda$ and $a_\lambda a_\nu = a_j K_{\lambda\nu}$ where $K_{\lambda\nu}$ is contained in K, and $\lambda, \nu = 1, \cdots, m$. Using the facts stated above, we finally give a necessary and a sufficient condition that the normal product $P = A\bar{A}$ be semi simple. We prove that P is semi simple if and only if A and $\bar{A}$ are both semi simple. In case the field F of the algebras A and $\bar{A}$ has a prime characteristic, some restrictions must be made for the truth of the above statement. The following notions and theorems will be often used:

1. If M is a finite module over an arbitrary field F, we denote by $|M|$ the order of M over F.[2]
2. The module M is said to be a *direct sum* of the modules $M_1, M_2, \cdots, M_r$ if: a) $M_\alpha \subset M$; b) If $m \subset M$, then $m = \sum_{\alpha=1}^{r} m_\alpha$, where $m_\alpha \subset M_\alpha$; c) From $\sum_{\alpha=1}^{r} m_\alpha = 0$, $m_\alpha \subset M_\alpha$ it follows $m_\alpha = 0$, $\alpha = 1, \cdots, r$. We write $M = \sum_{\alpha=1}^{r} M_\alpha$.
3. The *right* and *left ideals* and the *double ideals* of an algebra over F are defined as by E. Artin[3] or as in N, and are always modules over F.
4. A *primitive right ideal* R of an algebra A is a potent right ideal which does not contain potent right ideals of A except R itself. R possesses a left unit e and $R = eR = eA$. Every algebra A can

[2] For the definition of a module compare § 6 of E. Noether, Hyperkomplexe Größen und Darstellungstheorie, Math. Zeitschrift Bd. 30, 1929 p. 641–692. This paper will be cited henceforth with N.

[3] E. Artin, Zur Theorie der hyperkomplexen Zahlen, Abhandl. d. math. Sem. Universität Hamburg, **5** (1927) p. 251–260. Cited henceforth with A.

be represented in the form $A = \sum_{\lambda=1}^{s} e_\lambda A + N$, where the $e_\lambda A$ are primitive right ideals, and N is a nilpotent right ideal of A; further $A = eA + N$, where $eA = \sum_{\lambda=1}^{s} e_\lambda A$ and $e^2 = e$. The element e is called a *principal idempotent element* of A.

5. If R, R' are two sub-algebras of an algebra A, then we denote by $[R, R']$ the set of all the elements of A which are contained in R as well as in R'. This set is then also a sub-algebra of A.

By (R, R') we denote the minimal right ideal of A which contains R as well as R'. We say: (R, R') is *derived* from R and R'.

6. THEOREM. *Let A be an algebra over F, and M a potent sub-algebra of A over F. If M is the direct sum of the modules $M_1, M_2, \cdots, M_r$, then M contains a potent element p which has the form $p = p_1 \cdot p_2 \cdots p_t$, where each p_λ lies in an M_i.*

To prove this, let $m = \sum_{\alpha=1}^{r} m_\alpha$ where $m_\alpha \subset M_\alpha$ be a potent element of M. Since $0 \neq m^s = \sum m_{j_1} \cdot m_{j_2} \cdots m_{j_s}$ for every rational integer s, it follows that at least one of the elements $m_{j_1} \cdot m_{j_2} \cdots m_{j_s}$ is different from 0. If $s > |A|$, then it follows by a theorem of the writer[4] the existence of a potent element p which has the form $p = p_1 \cdot p_2 \cdots p_t$, where the p_i are certain elements of the system $m_{j_1}, \cdots, m_{j_s}$, and hence of $m_1, \cdots, m_r$, q. e. d.

7. Let $A = \sum_{\lambda=1}^{n} a_\lambda F$ be an algebra over F, and a an arbitrary element of A. Then $a_\lambda a = \sum_{\alpha=1}^{n} a_\alpha f_{\lambda\alpha}$ where $f_{\lambda\alpha} \subset F$. The correspondence $a \to (f_{\lambda\alpha})$ is called a *regular representation* of a, and the element $\sum_{i=1}^{n} f_{ii}$ of F the *regular trace* of a. Notation: $Tr_A(a) = \sum_{i=1}^{n} f_{ii}$. Traces are invariant under the transformations of the basis of A.

8. If a_λ, $\lambda = 1, \cdots, n$; b_ν, $\nu = 1, \cdots, n$ are two arbitrary basal systems of the algebra A over F, then the determinant $D_A = |Tr_A(a_\lambda \cdot b_\nu)|$ is called the *discriminant* of the algebra A, and is uniquely determined but for a factor from F which is different from zero; we shall also use the notation $D_A = D_A(a_\lambda \cdot b_\nu)$. If, (and in case the characteristic of F is 0 only if) $D_A \neq 0$, then the algebra A is semi simple.

9. Let $A = \bar{A} \times \bar{\bar{A}}$ be a direct product of two arbitrary algebras over an arbitrary field F. Then: a) If F has the characteristic 0, then A is semi simple if and only if both $\bar{A}$ and $\bar{\bar{A}}$ are semi simple. b) If F has a prime characteristic, then the condition of a) is necessary, and

[4] J. Levitzki, Über nilpotente Unterringe, § 2, Math. Ann. 105, p. 620–627.

in case the centres of $\bar{A}$ and $\bar{\bar{A}}$ do not contain inseparable fields, this condition is also sufficient in order that A be semi simple.

For the proof of this theorem see v. d. Waerden, Moderne Algebra II, § 119.

2. **The definition of a normal product.** Let $A = \sum_{i=1}^{m} a_i F$ and $\bar{A} = \sum_{k=1}^{\bar{m}} \bar{a}_k F$ be two algebras over an arbitrary field F, which are subsets of a given algebra. The product $P = A \cdot \bar{A}$ is called *normal with respect to* $\bar{A}$, if a basis $a_1, \cdots, a_m$ of A over F can be chosen such that $A \cdot \bar{A} = \sum_{\lambda} a_\lambda \bar{A}$ is a direct (right hand) module over $\bar{A}$ (i.e., from $\sum_{\lambda} a_\lambda \bar{b}_\lambda = 0$, where $\bar{b}_\lambda \subset \bar{A}$, it follows $\bar{b}_\lambda = 0$) and $a_\lambda \bar{A} = \bar{A} a_\lambda$ for $\lambda = 1, \cdots, m$.

The restriction, that A and $\bar{A}$ are subsets of a given algebra, can be replaced—analogous to the definition of a direct product—by the following set of axioms: We consider the set P of the linear forms $\sum_{\lambda=1}^{m} \sum_{\nu=1}^{\bar{m}} a_\lambda \bar{a}_\nu f_{\lambda\nu}$ and assume:

I. $a_\lambda \bar{a}_\nu f_{\lambda\nu} = f_{\lambda\nu} a_\lambda \bar{a}_\nu$ and $\left(\sum_{\lambda}\sum_{\nu} a_\lambda \bar{a}_\nu f_{\lambda\nu}\right) f = f\left(\sum_{\lambda}\sum_{\nu} a_\lambda \bar{a}_\nu f_{\lambda\nu}\right) = \sum_{\lambda}\sum_{\nu} a_\lambda \bar{a}_\nu (f_{\lambda\nu} f)$, where $f_{\lambda\nu}, f \subset F$.

II. If $a = \sum_{\lambda=1}^{m} a_\lambda f_\lambda$ resp. $\bar{a} = \sum_{\nu=1}^{\bar{m}} \bar{a}_\nu f_\nu$ are elements of A, resp. $\bar{A}$, then the element $a\bar{a}$ is defined by $a\bar{a} = \sum_{\lambda=1}^{m}\sum_{\nu=1}^{\bar{m}} a_\lambda \bar{a}_\nu f_{\lambda\nu}$, where $f_{\lambda\nu} = f_\lambda \cdot f_\nu$.

III. Two elements $p = \sum_{\lambda}\sum_{\nu} a_\lambda \bar{a}_\nu f_{\lambda\nu}$ and $p' = \sum_{\lambda}\sum_{\nu} a_\lambda \bar{a}_\nu f'_{\lambda\nu}$ of P are said to be equal, if and only if $f_{\lambda\nu} = f'_{\lambda\nu}$, $\lambda = 1, \cdots, m$; $\nu = 1, \cdots, \bar{m}$.

IV. The sum of two elements $p = \sum_{\lambda}\sum_{\nu} a_\lambda \bar{a}_\nu f_{\lambda\nu}$ and $p' = \sum_{\lambda}\sum_{\nu} a_\lambda \bar{a}_\nu f'_{\lambda\nu}$ of P is defined by $p + p' = \sum_{\lambda}\sum_{\nu} a_\lambda \bar{a}_\nu (f_{\lambda\nu} + f'_{\lambda\nu})$.

As consequences we obtain:

1. $(a+a')(\bar{a}+\bar{a}') = a\bar{a} + a\bar{a}' + a'\bar{a} + a'\bar{a}'$, where $a, a' \subset A$; $\bar{a}, \bar{a}' \subset \bar{A}$.
2. $(p+p')+p'' = p+(p'+p'')$ where $p, p', p'' \subset P$.
3. Every element of P can be written uniquely in the form $p = \sum_{\lambda=1}^{m} a_\lambda \bar{x}_\lambda$ where $\bar{x}_\lambda \subset \bar{A}$. We write $P = \sum_{\lambda=1}^{m} a_\lambda \bar{A} = A \cdot \bar{A}$.
4. If $a_1^*, \cdots, a_m^*$ resp. $\bar{a}_1^*, \cdots, \bar{a}_{\bar{m}}^*$ is an arbitrary basis of A, resp. $\bar{A}$, then the set P is identical with the set of all the elements of the form $\sum_{\lambda}\sum_{\nu} a_\lambda^* \bar{a}_\nu^* f_{\lambda\nu} = \sum_{\lambda}\sum_{\nu} a_\lambda^* (\bar{a}_\nu^* f_{\lambda\nu})$.

Vα. If $p = \sum_{\nu=1}^{m} a_\nu \bar{x}_\nu$ with $\bar{x}_\nu \subset \bar{A}$, and $\bar{a} \subset \bar{A}$, then $p\bar{a} = \sum_{\nu=1}^{m} a_\nu (\bar{x}_\nu \bar{a})$.

Vβ. If $p = \sum_{\nu=1}^{m} a_\nu \bar{x}_\nu$ with $\bar{x}_\nu \subset \bar{A}$, and $a \subset A$, then $ap = \sum_{\nu=1}^{m} (a a_\nu) \bar{x}_\nu$.

As consequences we obtain:

5. $(aa')p = a(a'p)$ where $a, a' \subset A$ and $p \subset P$.

6. $p(\bar{a}\bar{a}') = (p\bar{a})\bar{a}'$ where $\bar{a}, \bar{a}' \subset \bar{A}$ and $p \subset P$.

7. $(p+p')\bar{a} = p\bar{a}+p'\bar{a}$; $p(\bar{a}+\bar{a}') = p\bar{a}+p\bar{a}'$ where $p, p' \subset P$ and $\bar{a}, \bar{a}' \subset \bar{A}$.

8. $a(p+p') = ap+ap'$; $(a+a')p = ap+a'p$ where $p, p' \subset P$ and $a, a' \subset A$.

VI. $\bar{a}_\lambda a_\nu = a_\nu \bar{b}_{\lambda\nu}$ where $\bar{b}_{\lambda\nu} \subset \bar{A}$, $\lambda = 1, \cdots, m$; $\nu = 1, \cdots, \bar{m}$.

VII. If $\bar{a} = \sum_{\lambda} \bar{a}_\lambda \bar{f}_\lambda$ resp. $a = \sum a_\nu f_\nu$ are elements of $\bar{A}$, resp. of A, then we define $\bar{a}\,a$ by $\bar{a}\,a = \sum_{\nu} a_\nu \left(\sum_{\lambda} \bar{b}_{\lambda\nu} \bar{f}_\lambda f_\nu\right)$, where the $\bar{b}_{\lambda\nu}$ are defined by VI.

As consequences we obtain:

9. $(\bar{a}+\bar{a}')(a+a') = \bar{a}\,a+\bar{a}\,a'+\bar{a}'a+\bar{a}'a'$, where $\bar{a}, \bar{a}' \subset \bar{A}$; $a, a' \subset A$.

10. We can write VII in the form $\bar{A}\,a_\nu \subseteqq a_\nu \bar{A}$ and have $\sum_{\nu=1}^{m} \bar{A}\,a_\nu \subseteqq \sum_{\nu=1}^{m} a_\nu \bar{A} = P$.

We also write analogous to 3: $\sum_{\nu=1}^{m} \bar{A}\,a_\nu = \bar{A} \cdot A$.

Remark. While every equation of the form $\sum_{\nu=1}^{m} a_\nu \bar{x}_\nu = 0$, where $\bar{x}_\nu \subset \bar{A}$, according to 3 implies $\bar{x}_\nu = 0$, $\nu = 1, \cdots, m$, it does not follow that every equation of the form $\sum_{\nu=1}^{m} \bar{x}_\nu a_\nu = 0$ implies $\bar{x}_\nu = 0$, $\nu = 1, \cdots, m$.

VIII. $\bar{a}(\bar{a}'a) = (\bar{a}\bar{a}')a = \bar{a}\bar{a}'a$; $(\bar{a}a)a' = \bar{a}(aa') = \bar{a}aa'$ where $\bar{a}, \bar{a}' \subset \bar{A}$; $a, a' \subset A$.

IX. $\bar{a}\left(\sum_{\nu=1}^{m} a_\nu \bar{b}_\nu\right) = \sum_{\nu} (\bar{a}\,a_\nu)\bar{b}_\nu$; $\left(\sum_{\nu} a_\nu \bar{b}_\nu\right) a = \sum_{\nu} a_\nu (\bar{b}_\nu a)$ where $\bar{a}, \bar{b}_\nu \subset \bar{A}$; $a \subset A$.

Remark. Assumption IX defines the elements $\bar{a}\,p$ and $p\,a$, where $\bar{a} \subset \bar{A}$; $a \subset A$; $p \subset P$.

X. The product of two elements $p = \sum_{\nu=1}^{m} a_\nu \bar{b}_\nu$; $p' = \sum_{\nu=1}^{m} a_\nu \bar{b}'_\nu$ of P is defined as follows: $pp' = \sum_{\nu} \sum_{\mu} (a_\nu \bar{b}_\nu)(a_\mu b'_\mu) = \sum_{\nu} \sum_{\mu} (a_\nu \bar{b}_\nu a_\mu) \bar{b}'_\mu$.

As consequences we obtain:

11. $(p+p')(p''+p''') = pp''+pp'''+p'p''+p'p'''$ where $p, p', p'', p''' \subset P$.

12. $(pp')p'' = p(p'p'')$ where $p, p', p'' \subset P$.

13. The set P forms an algebra over F.

XI. Every relation of the form $\sum_{\nu=1}^{m} \bar{x}_\nu a_\nu = 0$ where $\bar{x}_\nu \subset \bar{A}$ implies $\bar{x}_\nu = 0$, $\nu = 1, \cdots, m$, (see remark to 10).

As consequences we obtain:

14. Instead of the relation $\bar{A} a_\lambda \subseteqq a_\lambda \bar{A}$ we have now the relation $\bar{A} a_\lambda = a_\lambda \bar{A}$ and hence $\bar{A}A = A\bar{A} = P$.

15. $(A\bar{A})^t = A^t \cdot \bar{A}^t = \bar{A}^t \cdot A^t$.

16. If $a_1, a_2, \cdots, a_t$ are elements of A which are independent over F, then they are also independent over $\bar{A}$, i. e.: From $a_1 \bar{b}_1 + \cdots + a_t \bar{b}_t = 0$ and $\bar{c}_1 a_1 + \cdots + \bar{c}_t a_t = 0$, where $\bar{b}_\lambda, \bar{c}_\nu \subset \bar{A}$, it follows $\bar{b}_\lambda = \bar{c}_\lambda = 0$, $\lambda = 1, \cdots, t$.

DEFINITION 1. The algebra P defined by I–XI is called a *normal product of the algebras A and $\bar{A}$ with respect to the algebra $\bar{A}$* (or in short: normal over $\bar{A}$). The algebra $\bar{A}$ is called a *normal factor of P*, and the algebra A a *supplementary factor of P with respect to $\bar{A}$* (or: over $\bar{A}$). The basis $a_1, \cdots, a_m$ of A which satisfies VI is called a *normal basis of A with respect to $\bar{A}$* (or: normal over $\bar{A}$).

DEFINITION 2. If, in particular, we assume in VI that $\bar{b}_{\lambda\nu} = \bar{a}_\nu$, $\lambda = 1, \cdots, m$; $\nu = 1, \cdots, \bar{m}$, then the normal product is called *direct*, and we write as usual $P = A \times \bar{A} = \bar{A} \times A$.

3. The automorphism-modules. DEFINITION 3. If $P = A\bar{A}$ is a normal product over $\bar{A}$, and x is a non zero element of A such that $\bar{A}x \subseteqq x\bar{A}$, then x is called a *normal element* of A over $\bar{A}$. From 2; 14 it follows then that $\bar{A}x = x\bar{A}$.

THEOREM 1. *Let $A\bar{A}$ be a normal product over $\bar{A}$ and x a normal element of A over $\bar{A}$. If $\bar{a}$ is an arbitrary element of $\bar{A}$, and $\bar{a}x = x\bar{b}$ where $\bar{b} \subset \bar{A}$, then the correspondence $\bar{a} \to \bar{b}$ is an automorphism of $\bar{A}$.*

To prove this, first note that to every element of $\bar{A}$ corresponds one and only one element of $\bar{A}$. Next, let $\bar{a}_1$ and $\bar{a}_2$ be two elements of $\bar{A}$ and $\bar{a}_1 \to \bar{b}_1$; $\bar{a}_2 \to \bar{b}_2$, i. e.: $\bar{a}_1 x = x\bar{b}_1$; $\bar{a}_2 x = x\bar{b}_2$; we obtain then: $(\bar{a}_1 + \bar{a}_2)x = \bar{a}_1 x + \bar{a}_2 x = x\bar{b}_1 + x\bar{b}_2 = x(\bar{b}_1 + \bar{b}_2)$ i. e., $\bar{a}_1 + \bar{a}_2 \to \bar{b}_1 + \bar{b}_2$. Further we have $(\bar{a}_1 \cdot \bar{a}_2)x = \bar{a}_1(\bar{a}_2 x) = \bar{a}_1(x\bar{b}_2) = (\bar{a}_1 x)\bar{b}_2 = (x\bar{b}_1)\bar{b}_2 = x(\bar{b}_1\bar{b}_2)$, hence $\bar{a}_1\bar{a}_2 \to \bar{b}_1\bar{b}_2$. The correspondence is therefore an homomorphism. But since $\bar{A}x = x\bar{A}$, this homomorphism is an automorphism, q. e. d.

DEFINITION 4. Automorphisms of the algebra $\bar{A}$ which are obtainable as the preceding theorem indicates, are said to be *induced by the algebra A* (or by its elements).

THEOREM 2. *If $P = A\bar{A}$ is a normal product over $\bar{A}$ and the elements a and b of A are normal over $\bar{A}$ and induce the same automorphism of $\bar{A}$, then*

also $af+bf'$ *(where* $f, f' \subset F$ *and* F *is the field of* A *and* $\bar{A}$*) is a normal element of* A *over* $\bar{A}$ *or zero, and it induces the same automorphism.*

If namely $\bar{x}$ is an arbitrary element of $\bar{A}$ then it follows from $\bar{x}a = a\bar{y}$; $\bar{x}b = b\bar{y}$, where $\bar{x}, \bar{y} \subset \bar{A}$ that $\bar{x}(af+bf') = (af+bf')\bar{y}$, q. e. d.

THEOREM 3. *If* $P = A\bar{A}$ *is a normal product over* $\bar{A}$ *and the elements* a *and* b *of* A *are normal over* $\bar{A}$*, and if their product is different from zero, then also the element* $a \cdot b$ *of* A *is normal over* $\bar{A}$ *and induces the automorphism* $\varphi \cdot \psi$ *of* $\bar{A}$*, where* φ, ψ *are respectively the automorphisms induced by* a *and* b.

To prove this, let $\bar{x}$ be an arbitrary element of $\bar{A}$. We have then $\bar{x}(ab) = (\bar{x}a)b = (a\bar{y})b = a(\bar{y}b) = a(b\bar{z}) = (ab)\bar{z}$, where $\bar{y}, \bar{z} \subset \bar{A}$. Hence the element $a \cdot b$ is normal over $\bar{A}$, and the automorphism $\bar{x} \to \bar{z}$ of $\bar{A}$ is the decomposition of φ and ψ, q. e. d.

THEOREM 4. *If* $P = A\bar{A}$ *is a normal product over* $\bar{A}$ *and the elements* $b_1, \cdots, b_s$, $s > 1$ *of* A *are normal over* $\bar{A}$ *and induce different automorphisms of* $\bar{A}$*, then*:

1. *The elements* $b_1, \cdots, b_s$ *are independent over* F*, where* F *is the field of* A *and* $\bar{A}$.
2. *Every element of the form* $\sum_{\lambda=1}^{s} b_\lambda f_\lambda$*, where* $f_\lambda \subset F$ *and more than one* f_λ *are different from zero, is not normal over* $\bar{A}$.

Suppose namely that the elements $b_1, \cdots, b_t$; $1 \leqq t < s$ are independent over F, while $b_{t+k} = \sum_{\lambda=1}^{t} b_\lambda f_{\lambda k}$; $f_{\lambda k} \subset F$; $k = 1, \cdots, s-t$. If $\bar{x}$ is an arbitrary element of $\bar{A}$, we have:

$$\bar{x}\, b_{t+k} = \bar{x} \sum_{\lambda=1}^{t} b_\lambda f_{\lambda k} = b_{t+k}\bar{x}' = \left(\sum_{\lambda=1}^{t} b_\lambda f_{\lambda k}\right)\bar{x}' = \sum_{\lambda=1}^{t} b_\lambda\, \bar{x}'\, f_{\lambda k}$$

where $\bar{x}' \subset \bar{A}$; on the other hand:

$$\bar{x} \sum_{\lambda=1}^{t} b_\lambda f_{\lambda k} = \sum_{\lambda=1}^{t} \bar{x}\, b_\lambda f_{\lambda k} = \sum_{\lambda=1}^{t} b_\lambda\, \bar{x}_\lambda\, f_{\lambda k}$$

where $\bar{x}_\lambda \subset \bar{A}$. Since the b_λ, $\lambda = 1, \cdots, t$ are independent over F, and hence over $\bar{A}$ (see 1;16) and since at least one $f_{\lambda k}$ is different from zero, we obtain for a certain λ the relation $\bar{x}' = \bar{x}_\lambda$. Hence the automorphism induced by b_{t+k} is equal to that induced by b_λ, in contradiction to the assumption that the b_ν induce different automorphisms. Hence $t = s$ and the first part of the theorem is proved.

To prove the second part, we write $c_\lambda = b_\lambda f_\lambda$ for those λ for which $f_\lambda \neq 0$, and may assume without loss of generality that $c_\lambda \neq 0$, $\lambda = 1, \cdots, t$;

$1 < t \leqq s$. Suppose the element $c = \sum_{\lambda=1}^{t} c_\lambda$ is normal over $\bar{A}$, i. e. for every $\bar{x} \subset \bar{A}$ we have

$$\bar{x}\, c = c\, \bar{x}' = \bar{x} \sum_{\lambda=1}^{t} c_\lambda = \left(\sum_{\lambda=1}^{t} c_\lambda\right) \bar{x}' = \sum_{\lambda=1}^{t} c_\lambda\, \bar{x}',$$

where $\bar{x}' \subset \bar{A}$; on the other hand

$$\bar{x} \sum_{\lambda=1}^{t} c_\lambda = \sum_{\lambda=1}^{t} \bar{x}\, c_\lambda = \sum_{\lambda=1}^{t} c_\lambda\, \bar{x}_\lambda,$$

where $\bar{x}_\lambda \subset \bar{A}$; since the c_λ are independent over F, and hence over $\bar{A}$, we have: $\bar{x}' = \bar{x}_1 = \cdots = \bar{x}_t$, i. e. the automorphisms induced by the c_λ are equal, in contradiction to the assumption. Hence c is not normal over $\bar{A}$, q. e. d.

As an immediate consequence of the first part of the preceding theorem, we obtain the following

THEOREM 5. *If $P = A\,\bar{A}$ is a normal product with respect to $\bar{A}$, then the number of the different automorphisms of $\bar{A}$ induced by A is finite, and does not exceed the order of the algebra A.*

To give a complete insight into the totality of normal elements of the algebra A over the normal factor $\bar{A}$, we prove

THEOREM 6. *If $P = A \cdot \bar{A}$ is a normal product over $\bar{A}$, and F is the field of A and $\bar{A}$, then A can be uniquely expressed as a direct sum of a finite number t of modules over the field F; $A = \sum_{\lambda=1}^{t} M_\lambda$, such that all the elements belonging to the same module are normal over $\bar{A}$ (or zero) and induce equal automorphisms of A, while elements which lie in different modules, induce different automorphisms. The number t is equal to the number of the different automorphisms of $\bar{A}$ induced by A, and every normal element of A over $\bar{A}$ lies in one of the modules M_λ.*

To prove this, we denote with $\psi_1, \psi_2, \cdots, \psi_t$ the different automorphisms of $\bar{A}$ which are induced by A. We denote by M_λ the set of all the normal elements of A over $\bar{A}$ which induce the automorphism ψ_λ of $\bar{A}$, inclusive the zero element. Since with two normal elements a_λ and b_λ of A which induce the automorphism ψ_λ also every element of the form $a_\lambda f + b_\lambda f'$, where $f, f' \subset F$ induces the automorphism ψ_λ or is equal to 0 (see Theorem 2), it follows that the set M_λ is a module over F. We thus obtain exactly t modules. If further b_λ, $\lambda = 1, \cdots, t$ are respectively elements of the modules M_λ, $\lambda = 1, \cdots, t$ which are different from zero, then we know by Theorem 4 that the b_λ are independent over F. Hence the sum of the modules M_λ is direct. From the definition of the modules M_λ it follows further that every normal element of A over $\bar{A}$ lies in one of

the M_λ, and since A possesses a normal basis over $\bar{A}$, it follows $A = \sum_{\lambda=1}^{t} M_\lambda$. The uniqueness of this representation is evident.

DEFINITION 5. The modules M_λ are called the *automorphism-modules* of the algebra A over the algebra $\bar{A}$.

REMARK. The preceding theorem gives a complete insight into the totality of the bases of A which are normal over $\bar{A}$. Every normal basis of A can be obtained by taking a set of t arbitrary bases of the t modules M_λ, and conversely, every basis of A thus obtained, is normal over $\bar{A}$.

4. **The kernel of the supplementary factor.** DEFINITION 6. If $P = A \cdot \bar{A}$ is a normal product over $\bar{A}$, then the subset K of all the elements of the supplementary factor A, which are commutative with every element of $\bar{A}$, is called the *kernel of A with respect to $\bar{A}$*, or over $\bar{A}$.

THEOREM 7. *If $P = A \cdot \bar{A}$ is a normal product over $\bar{A}$, then the kernel K of A over $\bar{A}$ is (in case $K \neq 0$) an algebra (over the field F of A and $\bar{A}$) and is identical with an automorphism-module of A over $\bar{A}$.*

The proof follows immediately from the definition of the kernel.

THEOREM 8. *Let $P = A \cdot \bar{A}$ be a normal product over $\bar{A}$, and let M_λ, $\lambda = 1, \cdots, t$, be the automorphism-modules of A over $\bar{A}$. If the product of two arbitrary modules M_α and M_β is not zero, then there exists one and only one module M_γ such that $M_\alpha M_\beta \subseteq M_\gamma$.*

If namely the elements of M_α induce the automorphism ψ_α of $\bar{A}$ and the elements of M_β induce the automorphism ψ_β, then every element of $M_\alpha M_\beta$ which is different from zero induces the automorphism $\psi_\alpha \psi_\beta$ of $\bar{A}$ (Theorem 3). Hence all the elements of $M_\alpha M_\beta$ belong to the same automorphism-module of A.

COROLLARY. If $M_1 = K$ is the kernel of A over $\bar{A}$, then $K M_\alpha \subseteq M_\alpha$ and $M_\alpha K \subseteq M_\alpha$ for $\alpha = 1, \cdots, t$.

THEOREM 9. *If A, $\bar{A}$ and the M_α, $\alpha = 1, \cdots, t$, are the same as in the preceding theorem, and b is a potent element of A which is normal over $\bar{A}$, then some power b^u of b, where $u \leqq t$, is commutative with every element of $\bar{A}$, i. e., $b^u \subset M_1$, where M_1 is the kernel of A over $\bar{A}$.*

We consider namely the powers $b, b^2, \cdots, b^{t+1}$ of b and denote with ψ the automorphism of $\bar{A}$ induced by b. The automorphism of $\bar{A}$ induced by b^λ is then by Theorem 3 equal to ψ^λ. Since t is the number of the different automorphisms of $\bar{A}$ induced by A, it follows that among the $t+1$ automorphisms ψ^λ, $\lambda = 1, \cdots, t+1$, at least two are identical: $\psi^j = \psi^k$, where $j < k \leqq t+1$; hence $(\psi^j)^{-1} \cdot \psi^j = (\psi^j)^{-1} \psi^k$, or $\psi^0 = \psi^{k-j}$, where ψ^0 denotes the identical automorphism of $\bar{A}$. Hence the element b^{k-j} induces the identical automorphism of $\bar{A}$, i. e. $b^{k-j} \bar{a} = \bar{a} b^{k-j}$ for every $\bar{a} \subset \bar{A}$, q. e. d.

COROLLARY. *If in the preceding theorem the element b is idempotent, then b is commutative with every element of $\bar{A}$. In this case we have namely (since $b^2 = b$) $k = 2$, $j = 1$, $k-j = 1$ and $b\bar{a} = \bar{a}b$ for every $\bar{a} \subset \bar{A}$.*

THEOREM 10. *If A, $\bar{A}$, M_α, $\alpha = 1, \cdots, t$, are the same as in the preceding theorem, then the supplementary factor A is a potent algebra if, and only if, the kernel $K = M_1$ of A over $\bar{A}$ is a potent algebra.*

If namely K is potent, then A is obviously potent. We now assume that A is potent. Since A is the direct sum of the modules M_α, it follows (see 1; 6) that there exists a potent element b in A which has the form $b = c_1 \cdot c_2 \cdots c_u$, where the c_i are elements of the modules M_α. According to Theorem 3, the potent element b is normal over $\bar{A}$. Hence (Theorem 9) some power of b, which is a potent element of A, belongs to the kernel of A, i. e., the kernel of A is potent, q. e. d.

5. The relationship between the supplementary factor and the kernel. Let $P = A\bar{A}$ be a normal product over $\bar{A}$ and let F be the field of A and $\bar{A}$. We then define:

DEFINITION 7. Every subset of A which forms a module over F, and whose basis over F can be chosen such that the elements (of the basis) are normal over $\bar{A}$, is called a *normal module over* $\bar{A}$. In particular we speak of sub-algebras (over F) of A which are normal over $\bar{A}$, and of normal right, left and (double) ideals of A over $\bar{A}$.

DEFINITION 8. The algebra A is called *normal primary over* $\bar{A}$, if A is a potent algebra, and does not contain any potent (double) ideal which is normal over $\bar{A}$ except A itself.

DEFINITION 9. *A potent right (left) ideal R of A* which is normal over $\bar{A}$ is called *normal primitive over* $\bar{A}$, if it does not contain potent right (left) ideals of A which are normal over $\bar{A}$, except A itself.

DEFINITION 10. The *algebra A* is called *normal primitive over* $\bar{A}$ if it does not contain neither right nor left potent ideals of A which are normal over $\bar{A}$ except A itself.

THEOREM 11. *Let $P = A \cdot \bar{A}$ be a normal product over $\bar{A}$ and $M^{(1)}$, $M^{(2)}$ two submodules of the algebra A over the field F of A and $\bar{A}$. If $M^{(1)}$ and $M^{(2)}$ are both normal over $\bar{A}$, and $M^{(1)} \cdot M^{(2)} \neq 0$, then also the module $M^{(1)} \cdot M^{(2)}$ over F is normal over F. In particular: If a is a normal element of A over $\bar{A}$, then $a \cdot A$ is normal over $\bar{A}$.*

We denote namely with $c_1^{(1)}, \cdots, c_s^{(1)}$ respectively $c_1^{(2)}, \cdots, c_t^{(2)}$ a basis of $M^{(1)}$ respectively of $M^{(2)}$ over F, which is normal over $\bar{A}$. Those among the elements $c_\lambda^{(1)} \cdot c_\nu^{(2)}$, $\lambda = 1, \cdots, s$; $\nu = 1, \cdots, t$ which are different from zero, are by Theorem 3 normal over $\bar{A}$. Choosing among the $c_\lambda^{(1)} \cdot c_\nu^{(2)}$ a set of linear independent elements over F, we obtain a normal basis of $M^{(1)} \cdot M^{(2)}$, q. e. d.

THEOREM 12. *If $P = A \cdot \bar{A}$ is a normal product over $\bar{A}$, then every normal primitive right ideal R of A possesses at least one left unit e (i. e. $R = eR = eA$) which is normal over $\bar{A}$. By corollary to Theorem 9, e is then an element of the kernel K of A over $\bar{A}$.*

To prove this, we consider the normal product $R \cdot \bar{A}$ over $\bar{A}$. The kernel K' of R over $\bar{A}$ is evidently equal to $[R, K]$, where K is the kernel of A over $\bar{A}$. Since R is potent, by Theorem 10 also K' is potent, i. e. there exists in K' a potent element a'. Since $a' \subset R$, we have $a'K \subseteqq RK \subseteqq R$; further, $a'K \subseteqq K$; hence $a'K \subseteqq [R, K]$. Let e be an idempotent element of the potent right ideal $a'K$ of K. Then $e \subset [R, K]$, hence $e \subset R$, and consequently $0 \subset eR \subseteqq R$; by Theorem 11 the potent right ideal eR of A is normal over $\bar{A}$, and since R is normal primitive over $\bar{A}$, it follows $eR = R$, i. e. the right ideal R possesses the left unit e which belongs to the kernel K of A, q. e. d.

COROLLARY. *If A is normal primitive over $\bar{A}$, then A possesses a unit element which belongs to the kernel of A, since in this case A possesses as well a right as a left unit, hence a unit e which is also the unit of the kernel K of A.*

THEOREM 13. *If $P = A\bar{A}$ is a normal product over $\bar{A}$, and K is the kernel of A over $\bar{A}$, then:*

1. *If $R = eA$ is an arbitrary normal primitive right ideal of A over $\bar{A}$ with the left unit e, where $e \subset K$ (see Theorem 12), then $[R, K]$ is a primitive right ideal of K, and $[R, K] = eK$.*
2. *If $R^* = e^*K$ is an arbitrary primitive right ideal of K with the left unit e^*, then the right ideal (R^*, R^*A) of A, derived from R^* and R^*A is normal primitive over $\bar{A}$, and $(R^*, R^*A) = e^*A$.*

To prove 1., first note that $[R, K]$ is a right ideal in K. Since $e \subset K$, we have $eK \subseteqq K^2 \subseteqq K$ and $eK \subseteqq eA \subseteqq R$, hence $eK \subseteqq [R, K]$; on the other hand, if k' is an arbitrary element of $[R, K]$, it is $ek' = k'$, since $k' \subset eR$; from $e \subset K$, $k' \subset K$ it follows $ek' = k' \subset eK$, i. e. $eK \supseteqq [R, K]$; hence $eK = [R, K]$. If now eK were not primitive, then $eK = e_1K + e_2K$ where $e_1^2 = e_1$, $e_2^2 = e_2$, $e_1e_2 = e_2e_1 = 0$. The right ideals e_1A and e_2A of A are by Theorem 11 normal over $\bar{A}$ and further $eA = e_1A + e_2A$. But this is a contradiction to the assumption that eA is normal primitive.

To prove 2. note that $e^*K \subseteqq e^*A$, hence $R^* \subseteqq R^*A$, hence $(R^*, R^*A) = R^*A$. From $R^*A \supseteqq e^*A$ and $R^*A = e^*(KA) \subseteqq e^*A$ we obtain $(R^*, R^*A) = e^*A$. Suppose e^*A were not normal primitive. Then let R be a normal primitive right ideal of A such that $R \subseteqq e^*A$ (such an ideal R exists). By Theorem 12 we can write $R = e'A$ where e' is idempotent in K. Since $e'A \supseteqq e'e^*A$, it is $e'e^* \neq e^*$. Writing $e^* = (e^* - e'e^*) + e'e^*$ and $e^* - e'e^* = e_1$, $e'e^* = e_2$ we obtain in the usual way the decomposition

$e^*K = e_1K + e_2K$ where $e_1^2 = e_1$, $e_2^2 = e_2$, $e_1e_2 = e_2e_1 = 0$, and $e_1K, e_2K \subset e^*K$. Hence e^*K is not primitive, in contradiction with the assumption.

THEOREM 14. *Let $P = A\bar{A}$ be a normal product over $\bar{A}$ and K the kernel of A over $\bar{A}$. Let $K = \sum_{i=1}^{u} e_iK + \bar{N}$ be a decomposition of K in a direct sum of primitive right ideals e_iK with the left hand units e_i, and the nilpotent right ideal $\bar{N}$. We further assume that $e = \sum_{i=1}^{u} e_i$ is a (principal) idempotent element of K. Then:*

1. *e is also a principal idempotent element of A.*
2. *It is $A = \sum_{i=1}^{u} e_iA + N$, where e_iA are normal primitive right ideals over $\bar{A}$, and N is a nilpotent right ideal of A which is normal over $\bar{A}$, and $N \supseteqq \bar{N}$.*

To prove 1., we consider the right ideal eA of A which contains the right ideal eK of K and which is normal over $\bar{A}$ (Theorem 11). Let a_λ, $\lambda = 1, \cdots, n$ be a normal basis of A over $\bar{A}$. An arbitrary system of linear independent elements (over F) among the $e \cdot a_\lambda$, $\lambda = 1, \cdots, n$ forms a normal basis of eA over $\bar{A}$, while a set of linear independent elements (over F) among the $ea_\lambda - a_\lambda$ forms a normal basis of a normal right ideal N of A, where $A = eA + N$. The elements $ea_\lambda - a_\lambda$ which are $\neq 0$ are namely normal over $\bar{A}$, since with a_λ also the element ea_λ induces the automorphism ψ_λ of $\bar{A}$ since e is commutative with every element of $\bar{A}$, and hence, by Theorem 2 also $ea_\lambda - a_\lambda$ is normal over $\bar{A}$ and induces the automorphism ψ_λ. Evidently $N \supseteqq \bar{N}$. Suppose the right ideal N of A is potent. By Theorem 9 N contains then a potent element b such that $b \subset K$. Since $e \subset K$ we obtain $K \supseteqq eK + bK$, where bK is a potent right ideal of K, i. e. e is not principal idempotent in K in contradiction to the assumption; hence the right ideal N of A is nilpotent and e is principal idempotent in A. The proof of 2. follows from the preceding theorem.

COROLLARY. *If in $P = A\bar{A}$ the algebra A has a unit e, then e is also the unit of the kernel K of A over $\bar{A}$, i. e.: The unit of A is commutative with every element of $\bar{A}$. If further $M_1, M_2, \cdots, M_t$ are the automorphism-modules of A over $\bar{A}$, and $K = M_1$, then $M_\alpha M_1 = M_1 M_\alpha = M_\alpha$ for $\alpha = 1, \cdots, t$.*

THEOREM 15. *If $P = A\bar{A}$ is a normal product over $\bar{A}$ and if A resp. $\bar{A}$ possesses the unit e, resp. $\bar{e}$, then $e\bar{e} = \bar{e}e = e^*$, and e^* is the unit of P. Further, P possesses two subalgebras: B, resp. $\bar{B}$ which are isomorphic to A, resp. $\bar{A}$, and each of which contains the unit e^*, and $P = B \cdot \bar{B}$, where $B \cdot \bar{B}$ is a normal product over $\bar{B}$.*

Let namely $a_1, \cdots, a_m$ be a normal basis of A over $\bar{A}$. For every $\bar{a}$ of $\bar{A}$ we have $\bar{a}\,a_\lambda = a_\lambda \bar{b}$, where $\bar{b} \subset \bar{A}$. Since the correspondence $\bar{a} \to \bar{b}$ is an automorphism of $\bar{A}$, $\bar{e}\,a_\lambda = a_\lambda \bar{e}$, for the unit is invariant under every automorphism. Hence the unit $\bar{e}$ of $\bar{A}$ is commutative with every element a of A. If further e is the unit of A, then by Theorem 14 (corollary) e is commutative with every element $\bar{a}$ of $\bar{A}$: $e\bar{a} = \bar{a}e$. The element $e^* = e\bar{e}$ is then the unit of P. If further a, resp. $\bar{a}$ is an arbitrary element of A resp. $\bar{A}$, then we write $b = a\bar{e} = \bar{e}a$; $\bar{b} = e\bar{a} = \bar{a}e$. The set of all the elements b, resp. $\bar{b}$ (which lie in P) form an algebra B, resp. $\bar{B}$ which lies in P and which is isomorphic to A resp. $\bar{A}$, and $P = B\bar{B}$ is a normal product over $\bar{B}$.

6. **Normal primary supplementary factors.** THEOREM 16. *If $P = A\bar{A}$ is a normal product over $\bar{A}$, then the supplementary factor A can be decomposed into a direct sum: $A = \sum_{\lambda=1}^{r} A_\lambda + N$ where the A_λ are normal primary algebras over $\bar{A}$, each containing a unit e_λ which belongs to the kernel K of A over $\bar{A}$, while N is a module over the field F of A and $\bar{A}$, which is normal over $\bar{A}$ and whose elements are properly nilpotent.*

Let namely $A = \sum d_i A + N'$ be a decomposition of A into the direct sum of normal primitive right ideals $d_i A$ and the normal nilpotent right ideal N', where $d_i^2 = d_i$; $d_i\,d_k = 0$, $i \neq k$ and where the d_i lie in the kernel of A over $\bar{A}$. Since in case $d_i A \cdot d_k A \neq 0$ this right ideal is (by Theorem 11) normal over $\bar{A}$, and since $d_i A \cdot d_k A \subseteq d_i A$, it follows that $d_i A \cdot d_k A$ is nilpotent (or zero), or $d_i A \cdot d_k A = d_i A$. We now classify the right ideals $d_i A$, making the following assumption: Two right ideals $d_i A$, $d_k A$ belong to the same class, if $d_i A \cdot d_k A = d_i A$. It follows easily that this definition satisfies the laws of symmetry, reflexivity and transitivity. Let R_λ denote the sum of those right ideals which belong to the λth class. Then $A = \sum_{\lambda=1}^{r} R_\lambda + N'$ where r denotes the number of the classes. The right ideal R_λ possesses a left unit e_λ, namely the sum of those d_i for which $d_i A \subseteq R_\lambda$. It is then $A = \sum_{\lambda=1}^{r} e_\lambda A + N'$ where $e_\lambda^2 = e_\lambda$, $e_\mu e_\nu = 0$, $\mu \neq \nu$ and $(e_\lambda A)^2 = e_\lambda A$, while $e_\lambda A \cdot e_{\lambda'} A$ nilpotent or zero if $\lambda \neq \lambda'$. If a_i denotes an arbitrary element of R_i, we write $a_i = a_i e_i + (a_i - a_i e_i)$ and obtain the decomposition $R_i = e_i A e_i + N_i$, where N_i belongs to the radical of A. Since e_i belongs to the kernel of A, it follows as in the proof of Theorem 14 that $e_i A e_i$ and N_i are normal over $\bar{A}$. We write now $A = \sum_{\lambda=1}^{r} A_\lambda + N$ where $A_\lambda = e_\lambda A e_\lambda$ and $N = \sum_{i=1}^{r} N_i + N'$.

It can be easily verified, that every (double) ideal of A_λ which is potent and normal over $\bar{A}$, is identical with A_λ, i. e. A_λ is normal primary over $\bar{A}$. This follows from the fact that A_λ possesses a unit element, and that for any two primitive right ideals $R_\lambda^{(1)}$, $R_\lambda^{(2)}$ of A_λ which are normal over $\bar{A}$ we have the relations $R_\lambda^{(1)} R_\lambda^{(2)} = R_\lambda^{(1)}$; $R_\lambda^{(2)} R_\lambda^{(1)} = R_\lambda^{(2)}$.

DEFINITION 11. The number r of the classes defined in the preceding theorem is called the *class number* of the normal primitive one hand ideals of A over $\bar{A}$. This number is an invariant of A with respect to $\bar{A}$.

DEFINITION 12. The number of the normal primitive one hand ideals whose sum is direct and which belong to the same class, is called the *normal length* of the class. This number is also an invariant of A with respect to $\bar{A}$.

THEOREM 17. *If $P = A\bar{A}$ is a normal product over $\bar{A}$, where A is normal primary over $\bar{A}$ and possesses a unit element, and if p is the normal length of A, then A contains a system of p^2 matric units $c_{\lambda\nu}$, $\lambda, \nu = 1, \cdots, p$ where $c_{\lambda\nu} \cdot c_{\nu'\mu} = \left.\begin{matrix} c_{\lambda\mu} \\ 0 \end{matrix}\right\}$ if $\begin{matrix} \nu = \nu' \\ \nu \neq \nu' \end{matrix}$ which are normal over $\bar{A}$. It is then $A = \sum c_{ii} A c_{kk}$, where the algebras $c_{ii} A c_{ii}$, $i = 1, \cdots, p$ are normal primitive over $\bar{A}$ and are isomorphic to each other. Finally, the automorphism of $\bar{A}$ induced by $c_{\lambda\nu}$ is inverse to that induced by $c_{\nu\lambda}$.*

First note, that the unit e of A is by the corollary to Theorem 14 commutative with every element of $\bar{A}$. Let further $A = \sum_{i=1}^{p} e_i A = \sum_{i=1}^{p} A e_i$ be a decomposition of A into normal primitive one hand ideals, where $e_\lambda^2 = e_\lambda$; $e_\lambda e_\nu = 0$; $\lambda \neq \nu$ and each e_λ is commutative with every element of $\bar{A}$. It is further $e_\lambda A e_\nu A = e_\lambda A$. This implies that the algebra $e_\lambda A e_\nu$ contains at least one element which is not properly nilpotent in A. If $a_i, i = 1, \cdots, n$ is a normal basis of A, then a set of linear independent elements among the $e_\lambda a_i e_\nu$ forms a normal basis of the normal algebra $e_\lambda A e_\nu$ over $\bar{A}$. At least one of the normal elements $e_\lambda a_i e_\nu$ is not properly nilpotent in A. Let then $c_{1\nu}$ be a non properly nilpotent element of $e_1 A e_\nu$ which is normal over $\bar{A}$. Since $c_{1\nu} A \subseteq e_1 A$, and the potent right ideal $c_{1\nu} A$ is normal over $\bar{A}$, it follows $c_{1\nu} A = e_1 A$, since the $e_i A$ are normal primitive. From $c_{1\nu} A = e_1 A$ it follows $c_{1\nu} \cdot e_\nu A e_1 = e_1 A e_1$; hence there exists an element $c_{\nu 1}$ in $e_\nu A e_1$ such that $c_{1\nu} c_{\nu 1} = e_1$. We shall prove that with $c_{1\nu}$ and e_1 also $c_{\nu 1}$ is normal over $\bar{A}$. To this end first note, that $c_{\nu 1} c_{1\nu} = e_\nu$. Next, let $\bar{a}$ be an arbitrary element of $\bar{A}$; then $c_{1\nu} \bar{a} = \bar{b} c_{1\nu}$, where $\bar{b} \subset \bar{A}$; hence $c_{\nu 1} c_{1\nu} \bar{a} c_{\nu 1} = c_{\nu 1} \bar{b} c_{1\nu} c_{\nu 1}$, or $e_\nu \bar{a} c_{\nu 1} = c_{\nu 1} \bar{b} e_1$; since e_1 and e_ν are commutative with every element of $\bar{A}$, we obtain $\bar{a} e_\nu c_{\nu 1} = c_{\nu 1} e_1 \bar{b}$ or $\bar{a} c_{\nu 1} = c_{\nu 1} \bar{b}$. With $\bar{a}$ also $\bar{b}$ runs over all the elements of $\bar{A}$, hence $c_{\nu 1}$ is normal over $\bar{A}$ and

implies evidently an automorphism of $\bar{A}$ which is the inverse of that implied by $c_{1\nu}$. Writing $c_{\lambda\nu} = c_{\lambda 1} \cdot c_{1\nu}$ we obtain a system of matric units $c_{\lambda\nu}$, where $c_{ii} = e_i$, which are by Theorem 3 normal over $\bar{A}$ and the automorphism of $c_{\lambda\nu}$ is inverse to that of $c_{\nu\lambda}$. The remaining part of the hypothesis follows easily from the relations $c_{\lambda\nu} A c_{\nu\lambda} = c_{\lambda\lambda} A c_{\lambda\lambda}$.

THEOREM 18. *If $P = A \cdot \bar{A}$ is a normal product over $\bar{A}$, where A is normal primary over $\bar{A}$ and possesses a unit element, then there exists a normal product $P' = A'\bar{A}'$, where $P = P'$, $A = A'$, and $\bar{A}'$ is a subalgebra of P which is isomorphic to $\bar{A}$. The algebra A can be decomposed into a direct product $A = C \times A^*$, where A^* is normal primitive over $\bar{A}'$, and C is a total matric algebra over the field F of A and $\bar{A}$, and the elements of C are commutative with every element of $\bar{A}'$, i. e.:*

$$P = A\bar{A}' = (C \times A^*) \cdot \bar{A}' = A^*(C \times \bar{A}') = C \times (A^* \cdot \bar{A}').$$

By Theorem 17 there exists in A a system of p^2 matric units $c_{\alpha\beta}$ which are normal over $\bar{A}$, such that $A = \sum_{\alpha,\beta} c_{\alpha\alpha} A c_{\beta\beta}$ where $c_{ii} A c_{ii}$ are normal primitive over $\bar{A}$, and each $c_{\alpha\alpha}$ is commutative with every element of $\bar{A}$. The subalgebras $c_{\alpha\alpha} \bar{A} c_{\alpha\alpha} = c_{\alpha\alpha} \bar{A}$ are isomorphic to each other and to $\bar{A}$, and $c_{\alpha 1} \bar{A} c_{1\alpha} = c_{\alpha\alpha} \bar{A} c_{\alpha\alpha}$. We consider the normal product $P_{\alpha\alpha} = c_{\alpha\alpha} A c_{\alpha\alpha} \cdot c_{\alpha\alpha} \bar{A} c_{\alpha\alpha}$ over the algebra $c_{\alpha\alpha} \bar{A} c_{\alpha\alpha}$. Since $c_{\beta\alpha} A c_{\alpha\beta} = c_{\beta\beta} A c_{\beta\beta}$, it follows that $c_{\beta\alpha} P_{\alpha\alpha} c_{\alpha\beta} = P_{\beta\beta}$, hence the algebras $P_{\beta\beta}$ are isomorphic to each other. Let now a be an arbitrary element of P_{11}. We then consider the element $p^* = \sum_{\alpha=1}^{p} c_{\alpha 1} a c_{1\lambda}$. The set of all the p^* forms an algebra P^* which is isomorphic to each $P_{\alpha\alpha}$, and the correspondence $p^* \to a$ is an isomorphism between P^* and P_{11}. Furthermore, it follows that the $c_{\alpha\beta}$ are commutative with every element of P^*, hence $P = C \times P^*$, where C is the total matric algebra of the $c_{\alpha\beta}$ over F. Every element p^* of P^* has the form $p^* = \sum_{\alpha=1}^{p} c_{\alpha 1} a c_{1\alpha}$ where $a \subset c_{11} A c_{11} \cdot c_{11} \bar{A} c_{11}$. Let a_{11}, resp. $\bar{a}_{11}$ be arbitrary elements in A_{11} resp. in $\bar{A}_{11}$. The element

$$\sum_{\alpha} c_{\alpha 1} a_{11} \bar{a}_{11} c_{1\alpha} = \sum_{\alpha} c_{\alpha 1} a_{11} c_{1\alpha} c_{\alpha 1} \bar{a}_{11} c_{1\alpha} = \sum_{\alpha} c_{\alpha 1} a_{11} c_{1\alpha} \cdot \sum_{\beta} c_{\beta 1} \bar{a}_{11} c_{1\beta}$$

is then an element of P^*. All the elements of the form $\sum_{\alpha} c_{\alpha 1} a_{11} c_{1\alpha}$ form an algebra A^* which is isomorphic to $c_{11} A c_{11}$ while all the elements of the form $\sum c_{\beta 1} \bar{a}_{11} c_{1\beta}$ form an algebra $\bar{A}'$ which is isomorphic to $c_{11} \bar{A} c_{11}$ and hence to $\bar{A}$, and it follows $P^* = A^* \bar{A}'$. From the construction of A^* and $\bar{A}'$ it follows further that $A^* \bar{A}'$ is a normal product over $\bar{A}'$, and that the $c_{\alpha\beta}$ are commutative with every element of A^* and $\bar{A}'$. Hence

$$P = C \times P^* = C \times (A^* \bar{A}') = (C \times A^*) \bar{A}' = A^*(C \times \bar{A}'), \text{ q. e. d.}$$

7. Normal primitive supplementary factors. THEOREM 19. *If $P = A\bar{A}$ is a normal product over $\bar{A}$, where A is normal primitive over $\bar{A}$, then A can be decomposed into a direct sum of two modules over the field F of A and $\bar{A}$, namely $A = T+N$, each being normal over $\bar{A}$. The module N is a subset of the radical of A, while the module T is an algebra over F which has the same unit as A, is normal primitive over $\bar{A}$, and each of its automorphism-modules over $\bar{A}$ contains at least one regular element. Finally, the automorphism-modules of T over $\bar{A}$ form a multiplicative group.*

We denote namely with M_α, $\alpha = 1, \cdots, t$ the automorphism-modules of A over $\bar{A}$. We denote further by $U_1, \cdots, U_g$ those M_α which do not contain any potent element, and by $V_1, \cdots, V_h$ the remaining M_α. Since A is normal primitive over $\bar{A}$, the modules U_i are subsets of the radical of A, while each of the modules V_k contains at least one regular element of A, since every element of V_k which is not nilpotent, is regular. Let namely a be a normal element of A over $\bar{A}$. The right ideal aA of A is by Theorem 11 normal over $\bar{A}$, and hence, since A is normal primitive, $aA = A$ or aA is nilpotent, i. e.: a is regular or a is properly nilpotent. It is $h+g = t$ where $0 < h \leqq t$. We write $N = \sum_{\alpha=1}^{g} U_\alpha$; $T = \sum_{\beta=1}^{h} V_\beta$, then $A = T+N$. According to Theorem 8, there exists a module M_γ such that $V_\alpha V_\beta \subseteq M_\gamma$. Since V_α and V_β contain regular elements, it follows that also M_γ contains regular elements, i. e. we can write $V_\alpha V_\beta \subseteq V_\gamma$. This shows that the module T is an algebra over F. It remains now to prove, that the modules V_α form a multiplicative group. We consider two arbitrary modules V_α and V_β. Let v_β be an arbitrary regular element of V_β. The element v_β^{-1} is then also normal over $\bar{A}$ and induces an automorphism of $\bar{A}$ which is reciprocal to that induced by v_β.[5] Hence the elements of the module $v_\beta^{-1} V_\beta$ imply the identical automorphism of $\bar{A}$ (i. e. they are commutative with every element of $\bar{A}$) and the module $v_\beta^{-1} V_\beta V_\alpha$ induces the same automorphism as V_α, hence $v_\beta^{-1} V_\beta V_\alpha \subseteq V_\alpha$. If v_α is an arbitrary regular element of V_α, we have in particular $v_\beta^{-1} V_\beta v_\alpha \subseteq V_\alpha$. Since v_α, v_β are regular, we have $|v_\beta^{-1} V_\beta v_\alpha| = |V_\beta|$ (where $|M|$ is the order of the module M over F). From $v_\beta^{-1} V_\beta v_\alpha \subseteq V_\alpha$ it follows $|V_\beta| \leqq |V_\alpha|$. Since the equation $|V_\beta| \leqq |V_\alpha|$ holds for $\alpha, \beta = 1, \cdots, h$, we obtain

[5] Let namely $\bar{a}$ be an arbitrary element of $\bar{A}$, then $\bar{a} v_\beta = v_\beta \bar{b}$, where $\bar{b} \subset \bar{A}$. Hence $v_\beta^{-1} \bar{a} v_\beta v_\beta^{-1} = v_\beta^{-1} v_\beta \bar{b} v_\beta^{-1}$ or $v_\beta^{-1} \bar{a} e = e \bar{b} v_\beta^{-1}$, where e is the unit of A. By Theorem 14, corollary, e is commutative with every element of $\bar{A}$, hence $v_\beta^{-1} e \bar{a} = \bar{b} e v_\beta^{-1}$, or $v_\beta^{-1} \bar{a} = \bar{b} v_\beta^{-1}$. Since with $\bar{a}$ also $\bar{b}$ runs over all the elements of $\bar{A}$, we have $\bar{A} v_\beta^{-1} = v_\beta^{-1} \bar{A}$, and it is also shown that the automorphism induced by v_β^{-1} is reciprocal to that induced by v_β.

$|V_\beta| = |V_\alpha|$, $\alpha, \beta = 1, \cdots, h$. From the equation $V_\alpha V_\beta \subseteqq V_\gamma$ it follows $|V_\alpha V_\beta| \leqq |V_\gamma|$; but since the V_i contain regular elements, we have evidently $|V_\alpha V_\beta| \geqq |V_\alpha| = |V_\gamma|$; hence $|V_\alpha V_\beta| = |V_\gamma|$ and this together with $V_\alpha V_\beta \subseteqq V_\gamma$ implies $V_\alpha V_\beta = V_\gamma$. Since from $V_\alpha V_\beta = V_\alpha V_{\beta'}$, it follows that $V_\beta = V_{\beta'}$, it is shown that the modules V_i form a group with the multiplication $V_\alpha V_\beta = V_\gamma$ as operation, q. e. d.

REMARK. The assumption, that every automorphism-module of the normal primitive supplementary factor contains at least one regular element, is equivalent with the assumption that the automorphism-modules form a group.

8. **The canonical representation of a normal product.** Combining the results of the previous paragraphs, we shall give in this paragraph a representation of a normal product, which we shall use often later on. We prove namely

THEOREM 20. *Let $P = A\bar{A}$ be a normal product over $\bar{A}$ and let F be the field of A, $\bar{A}$. Then $P = N\bar{A} + \sum_{\alpha=1}^{r} A_\alpha^* \bar{A}_\alpha$. Here is N a module over F which is normal over $\bar{A}$ and is a subset of the radical of A, while $A_\alpha^* \subseteqq A$, the $\bar{A}_\alpha$ are isomorph to each other and to $\bar{A}$, $\bar{A}_\alpha \subseteqq P$ and $A_\alpha^* \bar{A}_\alpha = A_\alpha^* \bar{A}$. Further, $A_\alpha^* A_\beta^* = 0$ if $\alpha \neq \beta$ and $A_\alpha^* = C_\alpha \times A_\alpha^{**}$, where C_α is a total matric algebra over F, and the elements of C_α are commutative with every element of $\bar{A}_\alpha$, while the A_α^{**} are normal primitive over $\bar{A}_\alpha$, so that $A_\alpha^* \bar{A}_\alpha = (C_\alpha \times A_\alpha^{**})\bar{A}_\alpha = C_\alpha \times (A_\alpha^{**} \bar{A}_\alpha) = A_\alpha^{**}(C_\alpha \times \bar{A}_\alpha)$. Finally, the automorphism-modules of A_α^* form a group.*

We represent namely the algebra A in the form $A = N' + \sum_{\alpha=1}^{r} A_\alpha$ where N' is a module over F, which is normal over $\bar{A}$ and is a subset of the radical of A, while the A_α are normal primary over $\bar{A}$ and contain unit elements (Theorem 16). We have then $A\bar{A} = N'\bar{A} + \sum_{\lambda=1}^{r} A_\lambda \bar{A}$. Each $A_\lambda \bar{A}$ we now represent by Theorem 18 in the form $A_\lambda \bar{A} = A_\lambda \bar{A}_\lambda$, where the $\bar{A}_\lambda$ are isomorphic to $\bar{A}$ (and hence also to each other) and $\bar{A}_\lambda \subseteqq A_\lambda \bar{A}$ (and hence $\bar{A}_\lambda \subseteqq P$); further $A_\lambda = C_\lambda \times A'_\lambda$, where C_λ is a total matric algebra over F, and A'_λ is normal primitive over $\bar{A}_\lambda$, so that

$$A_\lambda \bar{A}_\lambda = (C_\lambda \times A'_\lambda)\bar{A}_\lambda = C_\lambda \times (A'_\lambda \bar{A}_\lambda)$$

(see Theorem 18). Each A'_λ we represent in the form $A'_\lambda = A_\lambda^{**} + N'_\lambda$, where N'_λ is a module over F which is normal over $\bar{A}_\lambda$, and is a subset of the radical of A, while A_λ^{**} is normal primitive over $\bar{A}_\lambda$, and its automorphism-modules over $\bar{A}_\lambda$ form a group (Theorem 19). Hence

$$(C_\lambda \times A'_\lambda)\,\bar{A}_\lambda = [C_\lambda \times (A_\lambda^{**} + N'_\lambda)] \cdot \bar{A}_\lambda = [(C_\lambda \times A_\lambda^{**}) + (C_\lambda \times N'_\lambda)]\,\bar{A}_\lambda$$
$$= (C_\lambda \times A_\lambda^{**})\,\bar{A}_\lambda + (C_\lambda \times N'_\lambda)\,\bar{A}_\lambda .$$

Since N'_λ is a subset of the radical of A'_λ, it follows that $N_\lambda^* = C_\lambda \times N'_\lambda$ is a subset of the radical of A_λ, and hence of A. If c_{ik}^λ, $i, k = 1, \cdots, p_\lambda$ denotes the matric basis of C_λ over F, which was used to derive $\bar{A}_\lambda$ from $\bar{A}$ by the method of Theorem 19, then

$$N_\lambda^* = C_\lambda \times N'_\lambda = \sum_{i=1}^{p_\lambda} \sum_{k=1}^{p_\lambda} c_{i1}^\lambda\, N'_\lambda\, c_{1k}^\lambda \quad \text{and} \quad N_\lambda^*\,\bar{A}_\lambda = N_\lambda^*\,\bar{A},$$

where N_λ^* is also normal over $\bar{A}$. Writing

$$A_\alpha^* = C_\alpha \times A_\alpha^{**} \quad \text{and} \quad N = N' + \sum_{\lambda=1}^{r} N_\lambda^*$$

we obtain $P = N\bar{A} + \sum_{\lambda=1}^{r} A_\lambda^*\,\bar{A}_\lambda$, q. e. d.

9. **Group algebras modulo a sub-algebra.** By means of the canonical representation derived in the preceding paragraph, the study of the structure of a normal product $P = A \cdot \bar{A}$ is to some extent reduced to the case that A is normal primitive over $\bar{A}$ and its automorphism-modules over $\bar{A}$ form a group. If K is the kernel of A over $\bar{A}$, then the algebra A and its sub-algebra K are connected by very simple relations. We shall prove that A *is a group algebra modulo* K. This notion is defined as follows:

DEFINITION 13. The algebra A over the field F is said to be a *group algebra modulo its sub-algebra* A^* over the field F, if the following conditions are fulfilled:

I. A is a finite, direct, right hand module over A^*, i. e.: There exists a system of elements $a_1, \cdots, a_s$ in A, such that every element a of A can be uniquely expressed in the form $a = \sum_{\nu=1}^{s} a_\nu a_\nu^*$, where $a_\nu^* \subset A^*$. The a_λ are called the basis of A over A^*.

II. A is normal over A^*, i. e.: The basis a_λ can be chosen such that $a_\lambda A^* \subseteqq A^* a_\lambda$, $\lambda = 1, \cdots, s$.

REMARK. Since from $a_\lambda a^* = 0$ with $a^* \subset A^*$ it follows $a^* = 0$, the module $a_\lambda A^*$ has the same order over F as A^*: $|a_\lambda A^*| = |A^*|$. From $a_\lambda A^* \subseteqq A^* a_\lambda$ it follows $|a_\lambda A^*| \leqq |A^* a_\lambda|$. But evidently $|A^* a_\lambda| \leqq |A^*|$, hence $|a_\lambda A^*| = |A^* a_\lambda|$, and this together with $a_\lambda A^* \subseteqq A^* a_\lambda$ implies $a_\lambda A^* = A^* a_\lambda$. Hence, A is also a normal, direct left module over A^*.

III. To every pair of indices α, β, where $\alpha, \beta \leqq s$, there exists an index $\gamma \leqq s$ such that $a_\alpha A^* \cdot a_\beta A^* = a_\gamma A^*$.

REMARK. By remark to II we have also $A^* a_\alpha \cdot A^* a_\beta = A^* a_\gamma$.

IV. From $a_\alpha A^* \cdot a_\beta A^* = a_\alpha A^* \cdot a_{\beta'} A^*$ it follows $a_\beta A^* = a_{\beta'} A^*$, and from $a_\alpha A^* \cdot a_\beta A^* = a_{\alpha'} A^* \cdot a_\beta A^*$ it follows $a_\alpha A^* = a_{\alpha'} A^*$.

DEFINITION 14. The basis $a_1, \cdots, a_s$ of A over A^* which satisfies I-IV is called a *group basis of A modulo A^**. The group defined by III, IV is called the *group of A modulo A^**, and the number s is called the *order of A modulo A^**, or the order of the group of A modulo A^*.

THEOREM 21. a) *From Definition* 13, I–II *it follows that the algebra A^* possesses a unit element.*

b) *From Definition 13, I–IV it follows that also A possesses a unit element which is equal to that of A^*, and that the elements $a_1, \cdots, a_s$ (of a group basis of A modulo A^*) are regular in A.*

To prove a), we represent the a_λ in the form

$$a_\lambda = \sum_{\nu=1}^{s} a_\nu a^*_{\lambda\nu}, \quad a^*_{\lambda\nu} \subset A^*; \qquad \lambda = 1, \cdots, s. \tag{1}$$

If a^* is an arbitrary element of A^*, we obtain

$$a_\lambda a^* = \sum_{\nu=1}^{s} a_\nu a^*_{\lambda\nu} a^*. \tag{2}$$

Since the module $A = \sum_\nu a_\nu A^*$ is direct, we have

$$\begin{aligned} a^*_{\lambda\nu} a^* &= 0 \text{ for } \lambda \neq \nu, \\ a^*_{\lambda\lambda} a^* &= a^*. \end{aligned} \tag{3}$$

Hence, the elements $a^*_{\lambda\lambda}$, $\lambda = 1, \cdots, s$ are left units of A^*. Multiplying (1) from the left side with a^*, and using the fact that by II:

$$a^* a_\mu = a_\mu a^*_\mu, \text{ where } a^*_\mu \subset A^*, \tag{4}$$

we obtain

$$a^* a_\lambda = \sum_{\nu=1}^{s} a^* a_\nu a^*_{\lambda\nu}, \text{ or } a_\lambda a^*_\lambda = \sum_{\nu=1}^{s} a_\nu a^*_\nu a^*_{\lambda\nu}.$$

The module $A = \sum_{\nu=1}^{s} a_\nu A^*$ is direct, hence

$$\begin{aligned} a^*_\lambda a^*_{\lambda\nu} &= 0, \ \lambda \neq \nu \\ a^*_\lambda a^*_{\lambda\lambda} &= a^*_\lambda. \end{aligned} \tag{5}$$

The correspondence $a^* \to a_\lambda^*$ defined by (4) being an automorphism (the proof of this fact is the same as that of Theorem 1), it follows that with a^* also a_λ^* runs over all the elements of A^*, hence from (5) it follows that the $a_{\lambda\lambda}^*$, $\lambda = 1, \cdots, s$ are also right units of A^*, hence A^* possesses the unit $e^* = a_{11}^* = a_{22}^* = \cdots = a_{ss}^*$, and a) is proved.

COROLLARY: *If we replace in* (3) *the element* a^* *by* e^*, *we obtain* $a_{\lambda\nu}^* = 0$, $\lambda \neq \nu$.

To prove b), note that from II and the remark to II it follows $a_\lambda A^* \cdot a_\nu A^* = a_\lambda a_\nu (A^*)^2$. Since A^* possesses a unit, we have $(A^*)^2 = A^*$, and $a_\lambda A^* \cdot a_\nu A^* = a_\lambda a_\nu A^*$. For a fixed λ it follows from III–IV that for $\nu = 1, \cdots, s$ the module $a_\lambda A^* \cdot a_\nu A^*$ runs over all the $a_\mu A^*$, $\mu = 1, \cdots, s$. Hence

$$\sum_{\nu=1}^{s} a_\nu A^* = a_\lambda \sum_{\nu=1}^{s} a_\nu A^*, \text{ or } A = a_\lambda A.$$

Using the remark to III and the second part of IV, it follows similarly $A = A \cdot a_\lambda$. From $A = a_\lambda A$ it follows the existence of an element e such that $a_\lambda = a_\lambda e$. From $A = A a_\lambda$ it follows that e is a right unit of A. Similarly we prove the existence of a left unit e' of A. Hence $e = e'$ is the unit of A. From $A = A a_\lambda$ it follows the existence of an element $\bar{a}_\lambda$ such that $e = \bar{a}_\lambda a_\lambda$. Hence a_λ is regular in A and the element $\bar{a}_\lambda = a_\lambda^{-1}$ is reciprocal to a_λ. From the proof to a) and the corollary to that proof it follows that in (1) we have $a_\lambda = a_\lambda e^*$. But since a_λ is regular in A, it follows that the unit e^* of A^* is identical with the unit e of A, and this completes the proof of b).

THEOREM 22. *From Definition 13,* I–II *it follows, that if* A^* *has a radical* N^*, *then also* A *has a radical. It is* $a_\lambda N^* = N^* a_\lambda$, *and* $\sum_\lambda a_\lambda N^* = \sum_\lambda N^* a_\lambda = N$ *is a nilpotent* (*double*) *ideal in* A.

If namely a^* is an arbitrary element of A^*, then by II we have $a_\nu a^* = a_\nu^* a_\nu$ where $a_\nu^* \subset A^*$. The correspondence $a^* \to a_\nu^*$ is an automorphism of A^*. Since by every automorphism of an algebra, the radical is its own image we obtain

$$a_\nu N^* = N^* a_\nu \text{ and } \sum_\lambda a_\lambda N^* = \sum_\lambda N^* a_\lambda.$$

We obtain now

$$\left(\sum_\lambda a_\lambda A^*\right)\left(\sum_\nu a_\nu N^*\right) = \sum_\lambda \sum_\nu a_\lambda (A^* a_\nu N^*) = \sum_\lambda \sum_\nu a_\lambda (a_\nu A^*) N^*$$
$$= \left(\sum_\lambda \sum_\nu a_\lambda a_\nu A^*\right) N^* \subseteq \sum_\lambda a_\lambda (A^* N^*) \subseteq \sum_\lambda a_\lambda N^*,$$

i. e.,

$$\left(\sum_{\lambda} a_\lambda A^*\right)\left(\sum_{\nu} a_\nu N^*\right) \subseteq \sum_{\lambda} a_\lambda N^*;$$

similarly it follows that

$$\left(\sum_{\nu} a_\nu N^*\right)\left(\sum_{\lambda} a_\lambda A^*\right) \subseteq \sum_{\nu} a_\nu N^*,$$

hence $\sum_{\nu} a_\nu N^*$ is a (double) ideal in A. Further we obtain

$$\left(\sum_{\nu} a_\nu N^*\right)^2 = \sum_{\nu}\sum_{\lambda} a_\nu a_\lambda (N^*)^2 \subseteq \left(\sum_{\nu} a_\nu A^*\right)(N^*)^2 \subseteq \sum_{\nu} a_\nu (N^*)^2$$

and similarly for every t:

$$\left(\sum_{\nu} a_\nu N^*\right)^t \subseteq \sum_{\nu} a_\nu (N^*)^t.$$

If therefore, t is the exponent of the radical N^* of A^*, then $\left(\sum_{\nu} a_\nu N^*\right)^t = 0$, i. e., the ideal $\sum_{\nu} a_\nu N^*$ is nilpotent in A, q. e. d.

THEOREM 23. *If $A = \sum_{\nu=1}^{s} a_\nu A^*$ is a group algebra modulo A^*, and $a_1, a_2, \cdots, a_s$ is a group basis of A modulo A^*, then A^* is equal to a certain $a_\nu A^*$, and with $a_\nu A^*$, $\nu = 1, \cdots, s$ also the module $a_\nu^{-1} A^*$ runs over all the $a_\nu A^*$.*

This follows easily from the fact that the $a_\nu A^*$ form a group, and that the a_ν^* are by Theorem 21 regular in A.

LEMMA 1. *Let $A = \sum_{\nu=1}^{s} a_\nu A^*$ be a group algebra modulo A^*, where a_ν, $\nu = 1, \cdots, s$ is a group basis modulo A^*. Let a^* be an arbitrary element in A^*. Then $Tr_A(a^*) = s\, Tr_{A^*}(a^*)$* [6].

Let namely $a_1^*, \cdots, a_t^*$ be a basis of A^* over the field F of A^* (and A). Then $a_i a_k^*$, $i = 1, \cdots, s$; $k = 1, \cdots, t$ is a basis of A over F. Hence $a_k^* a^* = \sum_{j=1}^{t} \alpha_{jk} a_j^*$ with $\alpha_{jk} \subset F$, and we have

$$a_i a_k^* a^* = a_i \sum_{j=1}^{t} \alpha_{jk} a_j^* = \sum_{j=1}^{t} \alpha_{jk} (a_i a_j^*).$$

The matrix which represents a^* in the regular representation of A by the basis $a_i a_k^*$ consists, therefore, of s diagonal matrices, each of which being equal to the matrix (α_{jk}), $j, k = 1, \cdots, t$. This shows that $Tr_A(a^*) = s \sum_{j=1}^{t} \alpha_{jj}$. But $\sum_{j=1}^{t} \alpha_{jj} = Tr_{A^*}(a^*)$ hence $Tr_A(a^*) = s\, Tr_{A^*}(a^*)$, q. e. d.

[6] See 1; 7. Present lemma holds more generally also for such A and A^* which satisfy only the first condition of Definition 13.

LEMMA 2. *Let* $A = \sum_{\nu=1}^{s} a_\nu A^*$ *be the same as in Lemma 1. Then* $Tr_A(a_i a^* a_k^{-1} b^*) = 0$ *if* $i \neq k$. *Here* a_k^{-1} *is reciprocal to the regular element* a_k, *and* a^*, b^* *are arbitrary elements of* A^*.

Indeed $a_i a^* a_k^{-1} b^* = a_i a_k^{-1} c^*$, where $c^* \subset A^*$. From Theorem 23 it follows that there exists a certain j such that $a_i a_k^{-1} c^* = a_j d^*$, where $d^* \subset A^*$. Since $i \neq k$ it follows that a_j is not an element of A^* (i. e., $a_j A^* \neq A^*$). If now a_k^*, $k = 1, \cdots, t$ denotes an arbitrary basis of A^* over the field F of A^*, then $a_i a_k^*$, $i = 1, \cdots, s$; $k = 1, \cdots, t$ is a basis of A over F and $a_i a_k^* a_j d^* = a_\lambda p^*$ where $p^* \subset A^*$ and $a_\lambda \neq a_i$, since a_j is not an element of A^*. Hence, the matrix which corresponds to $a_j d^*$ in the regular representation of A by the basis $a_i a_k^*$; $i = 1, \cdots, s$; $k = 1, \cdots, t$ has zeros in its diagonal spaces, and this shows that $Tr_A(a_j d^*) = 0$, q. e. d.

LEMMA 3. *Let* A *be a group algebra modulo* A^*. *Then the bases of* A *and* A^* *over the field* F *can be chosen such that* $D_A = s^{ts}(D_{A^*})^s$ (*see* 1; 9) *where* t *is the order of* A^* *over* F, *and* s *is the order of the group of* A *modulo* A^*.

Let namely $a_1, \cdots, a_s$ be a group-basis of A modulo A^*, and $a_1^*, \cdots, a_t^*$ an arbitrary basis of A^* over F. Then the system $a_i^* a_k$; $i = 1, \cdots, t$; $k = 1, \cdots, s$ and by Theorem 23 also the system $a_i^{-1} a_k^*$; $i = 1, \cdots, s$; $k = 1, \cdots, t$ forms a basis of A over F. We write $u_{ik} = a_i^* a_k$ and $v_{ik} = a_i^{-1} a_k^*$. By Lemma 2 we know that if $k \neq \lambda$ then $Tr_A(u_{ik} v_{\lambda\nu}) = 0$. On the other hand

$$Tr_A(u_{ik} v_{k\nu}) = Tr_A(a_i^* a_k a_k^{-1} a_\nu^*) = Tr_A(a_i^* a_\nu^*);$$

by Lemma 2 we know that

$$Tr_A(a_i^* a_\nu^*) = s\, Tr_{A^*}(a_i^* a_\nu^*).$$

Hence the matrix $(Tr_A(u_{ik} v_{\lambda\nu}))$ consists of s diagonal matrices, each of which is equal to the matrix $(s\, Tr_{A^*}(a_i^* a_\nu^*))$ (while the remaining spaces of the matrix consist of zeros). The determinant of the matrix $(s\, Tr_{A^*}(a_i^* a_\nu^*))$ is $s^t D_{A^*}$. Since the discriminant of A is the determinant of $(Tr_A(u_{ik} v_{\lambda\nu}))$, we obtain $D_A = s^{ts}(D_{A^*})^s$, q. e. d.

COROLLARY. *If the order* s *of* A *modulo* A^* *is not divisible by the characteristic of the field* F *of* A *and* A^*, *then the discriminant of* A *is zero, if and only if the discriminant of* A^* *is zero. From* 1; 8 *it follows now*

THEOREM 24. *If the algebra* A *over* F *is a group algebra modulo the sub-algebra* A^* *over* F, *where the field* F *has the characteristic zero, then the algebra* A *is semi simple, if, and only if the algebra* A^* *is semi simple.*

Returning to the normal products we now prove

THEOREM 25. *If $P = A\bar{A}$ is a normal product over $\bar{A}$, and A is normal primitive over $\bar{A}$ and its automorphism-modules over $\bar{A}$ form a group, then A is a group algebra modulo the sub-algebra K of A, where K is the kernel of A over $\bar{A}$.*

Let namely $M_1, \cdots, M_t$ be the automorphism modules of A over $\bar{A}$, and let $M_1 = K$ be the kernel of A; then $M_\lambda M_1 = M_1 M_\lambda = M_\lambda$ for every $\lambda = 1, \cdots, t$. Let a_λ be a regular element of M_λ (see the remark to Theorem 19). Then $a_\lambda M_1 \subseteq M_\lambda$. But since a_λ is regular, $|a_\lambda M_1| = |M_1|$, and this is by the proof to Theorem 19, also equal to $|M_\lambda|$. Consequently, $a_\lambda M_1 = M_\lambda$ and hence $A = \sum_\lambda M_\lambda = \sum_\lambda a_\lambda M_1$. Similarly it follows that $M_1 = M_1 a_\lambda$, i. e. $a_\lambda M_1 = M_1 a_\lambda$. Since the sum of the M_λ is direct, the postulates I–II of Definition 13 are fulfilled. Since, further, the M_λ form a group, also the postulates III–IV of Definition 13 are fulfilled. Hence A is a group algebra modulo M_1, q. e. d.

THEOREM 26. *If $P = A\bar{A}$ is a normal product over $\bar{A}$ and A is a group algebra modulo its kernel K over $\bar{A}$, and if further $\bar{A}$ possesses a unit element (by Theorem 21 the algebra A has a unit), then we can replace A and $\bar{A}$ by isomorphic algebras B, resp. $\bar{B}$, such that $P = B \cdot \bar{B}$ is a normal product over $\bar{B}$, and $B, \bar{B} \subset P$. Further, B is a group algebra modulo its kernel K^* over $\bar{B}$, the algebra P is a group algebra modulo the sub-algebra $K^* \times \bar{B}$, and the groups of A modulo K, of B modulo K^*, and of P modulo $K^* \times \bar{B}$ are isomorphic to each other.*

Let namely $a_1, \cdots, a_s$ be a group basis of A modulo K, chosen as indicated in the proof of Theorem 25. The a_λ, $\lambda = 1, \cdots, s$ are then normal over $\bar{A}$. We replace now the algebras A and $\bar{A}$ by the method of Theorem 15 by isomorphic sub-algebras B, resp. $\bar{B}$ of P. If by this replacement the elements $b_1, \cdots, b_s$ correspond to the elements $a_1, \cdots, a_s$ of A, and K^* corresponds to K, then the b_i are normal over $\bar{B}$ and form a group-basis of B modulo K^*, and K^* is the kernel of B over $\bar{B}$. Then

$$P = B\bar{B} = \left(\sum_{\lambda=1}^{s} b_\lambda K^*\right)\bar{B} = \sum_{\lambda=1}^{s} b_\lambda (K^* \times \bar{B}),$$

and these relations complete the proof of the theorem, since

$$\begin{aligned}[b_\lambda (K^* \times \bar{B})] \cdot [b_\nu (K^* \times \bar{B})] &= b_\lambda (K^* \times \bar{B}) \cdot b_\nu (K^* \times \bar{B}) \\ &= (b_\lambda b_\nu K^*)\bar{B} = b_\gamma (K^* \times \bar{B})\end{aligned}$$

where $b_\lambda b_\nu K^* = b_\gamma K^*$, in correspondence to $a_\lambda a_\nu K = a_\gamma K$ or $(a_\lambda K)(a_\nu K) = a_\gamma K$.

10. Necessary and sufficient conditions that a normal product be semi simple. LEMMA 1. *Let $P = A\bar{A}$ be a normal product over the algebra $\bar{A}$. If $\bar{A}$ has the radical $\bar{N}$, then $N^* = A\bar{N} = \bar{N}A$, and N^* is a normal product over $\bar{N}$ and is a nilpotent ideal of P, i. e.: In order that P be semi simple, it is necessary that $\bar{A}$ be semi simple.*

Let namely $a_1, \cdots, a_m$ be a normal basis of A with respect to $\bar{A}$. If $\bar{a} \subset \bar{A}$, then $\bar{a}\, a_\lambda = \bar{a}_\lambda\, a_\lambda$ where $\bar{a}_\lambda \subset \bar{A}$. Since the correspondence $\bar{a} \to \bar{a}_\lambda$ is an automorphism of $\bar{A}$, and by every automorphism of $\bar{A}$ the radical $\bar{N}$ of $\bar{A}$ is its own image, it follows that $\bar{N} a_\lambda = a_\lambda \bar{N}$, and hence $A\bar{N} = \bar{N}A$. Further we obtain $A\bar{N} \cdot A\bar{A} = AA \cdot \bar{N}\bar{A}$ and since $A^2 \subseteq A$, $\bar{N}\bar{A} \subseteq \bar{N}$, we have $A\bar{N} \cdot P \subseteq A\bar{N}$. Similarly $P \cdot A\bar{N} \subseteq A\bar{N}$. Further we have $(A\bar{N})^\lambda = A^\lambda \cdot \bar{N}^\lambda$. If then q is the exponent of the radical $\bar{N}$ of $\bar{A}$, we have $(A\bar{N})^q = 0$, and this completes the proof of the theorem.

LEMMA 2. *Let $P = A\bar{A}$ be a normal product over the algebra $\bar{A}$. If a is a normal element of A over $\bar{A}$, which is properly nilpotent in A, then the elements of the module $a\bar{A}$ are properly nilpotent in P, i. e.: In order that P be semi simple, it is necessary that A does not possess normal elements over $\bar{A}$ which are properly nilpotent in A.*

It is namely $(a\bar{A})\, P = a\bar{A} \cdot A\bar{A} = aA \cdot \bar{A}^2$; since a is properly nilpotent in A, it follows that the algebra aA (in case $aA \neq 0$) is a right hand nilpotent ideal of A which is normal over $\bar{A}$ (this follows from Theorem 11). Hence $(aA\bar{A}^2)^\lambda = (aA)^\lambda \cdot \bar{A}^{2\lambda}$. If q is the exponent of the nilpotent right hand ideal aA of A, then $(aA \cdot \bar{A}^2)^q = (a\bar{A} \cdot P)^q = 0$, i. e. the module $a\bar{A}$ is a properly nilpotent subset of P, q. e. d.

THEOREM 27. *If $P = A \cdot \bar{A}$ is a normal product over the algebra $\bar{A}$, and the field F of the algebras A and $\bar{A}$ has the characteristic zero,*[7] *then a necessary and a sufficient condition that the algebra P be semi simple is, that the algebras A and $\bar{A}$ are both semi simple.*

If namely $\bar{A}$ has a radical, then by Lemma 1 P has also a radical. We now assume that $\bar{A}$ is semi simple and A has a radical. Using the notations of Theorem 20 we consider a canonical representation of P: $P = N\bar{A} + \sum_\alpha A_\alpha^* \bar{A}_\alpha$ where the $\bar{A}_\alpha$ are semi simple (since they are isomorphic to $\bar{A}$). If $N \neq 0$, then by Lemma 2 the algebra P has a radical. If $N = 0$ then (since $A_\alpha^* A_\beta^* = 0$ for $\alpha \neq \beta$) at least one of the algebras A_α^* must have a radical. Further $A_\alpha^* = C_\alpha \times A_\alpha^{**}$. Since C_α is a total matric algebra over F, it follows by 1; 9 that A_α^{**} has a radical. Since A_α^{**} and $\bar{A}_\alpha$ have unit elements, we can replace them by isomorphic sub-algebras B_α^{**}, $\bar{B}_\alpha$ of $A_\alpha^{**} \cdot \bar{A}_\alpha$ such that $A_\alpha^{**} \bar{A}_\alpha = B_\alpha^{**} \cdot \bar{B}_\alpha$ is a normal product over $\bar{B}_\alpha$ (see Theorem 18). The algebra B_α^{**}, being isomorphic to A_α^{**}, has a radical.

[7] The case of a prime characteristic is treated in the next paragraph.

If K_α is the kernel of B_α^{**} over $\overline{B}_\alpha$, then by Theorems 24 and 25 the algebra K_α has a radical. Applying 1; 9 it follows that $K_\alpha \times \overline{B}_\alpha$ has a radical. Since $B_\alpha^{**} \times \overline{B}_\alpha$ is a group algebra modulo $K_\alpha \times \overline{B}_\alpha$ (see Theorem 26) it follows by Theorem 24 that also $B_\alpha \times \overline{B}_\alpha$ has a radical. Since further $A_\alpha^* \overline{A}_\alpha = C_\alpha \times (B_\alpha^{**} \overline{B}_\alpha)$, it follows by 1; 9, that $A_\alpha^* \overline{A}_\alpha$, and hence P has a radical. Conversely, we now assume that A and $\overline{A}$ are without radical. Since in this case A and $\overline{A}$ possess unit elements, we can replace them by isomorphic sub-algebras of P (Theorem 18). We therefore assume that in $P = A\overline{A}$ the algebras A and $\overline{A}$ are sub-algebras of P. In the canonical representation of P we have now $N = 0$, while the A_α^* are semi simple. Since $A_\alpha^* = C_\alpha \times A_\alpha^{**}$, it follows (1; 9) that A_α^{**} is semi simple, and hence (Theorems 24, 25), that also the kernel K_α of A_α^{**} over $\overline{A}_\alpha$ is semi simple. By 1; 9, $K_\alpha \times \overline{A}_\alpha$ is semi simple, and hence (Theorem 26) also $A_\alpha^{**} \overline{A}_\alpha$. By 1; 9 $A_\alpha^* \overline{A}_\alpha = C_\alpha \times (A_\alpha \overline{A}_\alpha)$ is semi simple, and consequently also the algebra $P = \sum_\alpha A_\alpha^* \overline{A}_\alpha$, since $(A_\alpha^{**} \overline{A}_\alpha)(A_\beta^{**} \overline{A}_\beta) = 0$, if $\alpha \neq \beta$. This completes the proof of the theorem.

APPENDIX: **The case of a prime characteristic.** LEMMA 1. *Let A be a group algebra over the sub-algebra A^*, where the field F of A and A^* has a prime characteristic, and A^* is a semi simple total matric algebra over F. Then A is semi simple if the characteristic of F is not smaller than the order of A.*

We write $A^* = \sum_{i=1}^{r} A_i^*$, where A_i^* are simple total matric algebras over F. Let c_{ik}^λ, $i, k = 1, \cdots, n_\lambda$ be a matric basis of A_λ^* over F. Then each of the systems $c_{ik}^\lambda a_\nu$ and $a_\nu^{-1} \overline{c}_{ik}^\lambda$, (where a_ν, $\nu = 1, \cdots, s$ is a group basis of A over A^*, and $\overline{c}_{ik}^\lambda = c_{ki}^\lambda$), forms a basis of A over F. As in Lemma 3 of 9 we obtain $D_A = s^{ts} D_{A^*}(c_{ik}^\lambda \cdot \overline{c}_{\alpha\beta}^\gamma)$, where s is the order of the group of A modulo A^*, and t is the order of the algebra A^*, i. e., $t = \sum_{\lambda=1}^{r} n_\lambda^2$. Using the matric-units-relations between the c_{ik}^λ we obtain $D_{A^*}(c_{ik}^\lambda \overline{c}_{\alpha\beta}^\gamma) = n_1^{n_1^2} \cdots n_r^{n_r^2}$. (See N, § 25). Hence $D_A = s^{ts} \cdot n_1^{n_1^2} \cdots n_r^{n_r^2}$. Since s, n_λ are smaller than the order of A, it follows $D_A \neq 0$. By 1; 8 the algebra A is semi simple.

LEMMA 2. *If A is a group algebra modulo the sub-algebra A^*, where the field F of A and A^* has a prime characteristic which is not smaller than the order of A, and if the centre of A^* does not contain inseparable extensions of F, then A is semi simple if, and only if A^* is semi simple.*

If namely A^* has a radical, then by Theorem 22 also A has a radical. We now assume that A^* is semi simple. We extend the field F of A^* to an algebraic closed field F^*, and consider the algebra $A^* \times F^*$. By 1; 9,

the algebra $A^* \times F^*$ is semi simple, and is therefore a direct sum of simple total matric algebras over F^*. We consider now the algebra $A \times F^* = \left(\sum_{\lambda=1}^{s} a_\lambda A^*\right) \times F^* = \sum_{\lambda=1}^{s} a_\lambda (A^* \times F^*)$, where the a_λ form a group basis of A over A^*. The algebra $A \times F^*$ is then a group algebra modulo the total semi simple matric algebra $A^* \times F^*$. By a properly chosen basis of $A \times F^*$ over F^* we obtain as in Lemma 1: $D_{A \times F^*} = s^{ts} \cdot n_1^{n_1^2} \cdots n_r^{n_r^2}$, where s is the order of the group of A modulo A^*, the n_λ^2 are the orders of the simple ideals of $A^* \times F^*$ and t is the order of A^* over F (or of $A^* \times F^*$ over F^*). Since $s, n_1, \cdots, n_r$ do not exceed the order of A over F, we obtain $D_{A \times F^*} \neq 0$, hence, the algebra $A \times F^*$ is semi simple, and consequently also A.

Using the facts stated in this paragraph and the same conclusions as in the proof of Theorem 27, we now obtain

THEOREM 27a. *Let $P = A\bar{A}$ be a normal product over $\bar{A}$, where the characteristic of the field F of A and $\bar{A}$ is not smaller than the order of P over F. If further the centres of A and $\bar{A}$ do not contain fields which are inseparable extensions of F, then a necessary and a sufficient condition that P be semi simple, is that A and $\bar{A}$ are both semi simple.*

DEFINITION OF A FIELD BY FOUR POSTULATES.[1]

By Wallie Abraham Hurwitz.

1. **Introduction.** Investigations of the definition of a field by independent postulates were made by L. E. Dickson,[2] E. V. Huntington,[3] the writer,[4] and N. Wiener.[5] Each of these studies sought to simplify the formal logical structure of the definition in one way or another. One of Huntington's postulate-sets did not assume the law for closure with respect to addition or multiplication, but in each case proved it. I combined the commutative and associative laws of addition into a single postulate from which those laws followed, and similarly for multiplication. Wiener, taking a different direction of attack, introduced addition and multiplication, not as undefined concepts appearing in the postulates, but as consequences of a distinct single law of combination; this procedure was analogous to that of H. M. Sheffer[6] in the definition of a boolean algebra.

In each of these cases, the authors proved that the postulate-set given was consistent and independent, sometimes even with the inclusion of a postulate restricting the cardinal number (e. g., finite, infinite, having a stated finite number of elements, countably infinite, etc.). None of them considered the question of complete independence, as defined by E. H. Moore.[7] However, B. A. Bernstein[8] later studied exhaustively the complete existential theory of my set of postulates.

The least numbers of postulates used are seven in one set of Huntington and that of Wiener, and five in my own set. In this paper I make another step in the direction of my earlier work, retaining the combination of commutative and associative laws for addition, but consolidating into a single postulate the commutative and associative laws for multiplication and the distributive law. The resulting set of four postulates is shown to be completely independent, even with the inclusion of a postulate specifying finiteness or infiniteness.

[1] Received September 4, 1931.—Presented to the American Mathematical Society.

[2] Transactions of the American Mathematical Society, vol. 4 (1903), pp. 13–26, and vol. 6 (1905), pp. 198–208; Gött. Nachr., Mathematisch-Physikalische Klasse (1905), pp. 358–393.

[3] Transactions of the American Mathematical Society, vol. 4 (1903), pp. 31–37, and vol. 6 (1905), pp. 181–197.

[4] These Annals, second series, vol. 15 (1913–14), pp. 93–100. This paper, which dealt with both abelian groups and fields, will be cited as AGF.

[5] Transactions of the American Mathematical Society, vol. 21 (1920), pp. 237–246.

[6] Transactions of the American Mathematical Society, vol. 14 (1913), pp. 481–488.

[7] Introduction to a form of general analysis, New Haven Mathematical Colloquium, 1910, p. 82. Of course, Moore's work is subsequent to the cited papers of Dickson and Huntington.

[8] These Annals, second series, vol. 23 (1921–22), pp. 313–316; Bulletin of the American Mathematical Society, vol. 28 (1922), pp. 397–399.

2. **The postulates.** The undefined notions are a set K of elements and two combinations, each uniquely determined, of pairs of elements, $a \oplus b$ and $a \otimes b$, satisfying the following conditions:

($A1$) *If* $a, b, c, a \oplus b, c \oplus b$, *and* $a \oplus (c \oplus b)$ *belong to* K, *then* $(a \oplus b) \oplus c$ *belongs to* K *and*

$$(a \oplus b) \oplus c = a \oplus (c \oplus b).$$

($A2$) *If* a *and* b *belong to* K, *there is an element* x *of* K *such that*

$$a \oplus x = b.$$

(M) *If* a *and* b *belong to* K *and* $a \oplus a \neq a$, *there is an element* x *of* K *such that*

$$a \otimes x = b.$$

(S) *If* $a, b, c, d, a \otimes b, a \otimes d, c \otimes d, b \oplus (c \otimes d)$, *and* $a \otimes \{b \oplus (c \otimes d)\}$ *belong to* K, *then* $(a \otimes d) \otimes c$ *and* $(a \otimes b) \oplus \{(a \otimes d) \otimes c\}$ *belong to* K, *and*

$$a \otimes \{b \oplus (c \otimes d)\} = (a \otimes b) \oplus \{(a \otimes d) \otimes c\}.$$

The cardinal number will be restricted by one of the following postulates:

(Nf) K *is finite with at least two elements.*

($N\infty$) K *is infinite.*

3. **Deductions from the postulates.** The postulates ($A1$), ($A2$) are identical with (1), (2) in § 1 of AGF with $\oplus$ for 0; hence

I. K *is an abelian group with respect to* $\oplus$.

This result implies among other things that for every a, b in K, $a \oplus b$ is in K; and that there is a unique identity 0 in K such that for every a in K, $a \oplus 0 = 0 \oplus a = a$; for no $a \neq 0$ is $a \oplus a = a$.

II. *If* a *is in* K, *then* $0 \otimes a = 0$.

Suppose first that $a \neq 0$. By (M) choose b so that $a \otimes b = 0$; apply (S) with $c = a$, $d = b$:

$$a \otimes \{b \oplus (a \otimes b)\} = (a \otimes b) \oplus \{(a \otimes b) \otimes a\},$$

or

$$0 = 0 \oplus (0 \otimes a);$$

hence, by I, $0 \otimes a = 0$.

Now let $a = 0$; we wish to prove that $0 \otimes 0 = 0$. In (S) put $c = a = 0$, $d = b \neq 0$; by the preceding proof $0 \otimes b = 0$:

$$0 \otimes \{b \oplus (0 \otimes b)\} = (0 \otimes b) \oplus \{(0 \otimes b) \otimes 0\},$$

or

$$0 = 0 \oplus (0 \otimes 0);$$

hence, by I, $0 \otimes 0 = 0$.

III. *If* a *is in* K, *then* $a \otimes 0 = 0$.

Let $c \neq 0$; by (M) take b so that $c \otimes b = a$; apply (S), replacing a, b, c, d respectively by $c, b, 0, b$:

$$c \otimes \{b \oplus (0 \otimes b)\} = (c \otimes b) \oplus \{(c \otimes b) \otimes 0\},$$

or, by II,

$$a = a \oplus (a \otimes 0);$$

hence, by I, $a \otimes 0 = 0$.

IV. *If* $a, b, c, a \otimes b, c \otimes b,$ *and* $a \otimes (c \otimes b)$ *are in* K, *then* $(a \otimes b) \otimes c$ *is in* K, *and*

$$(a \otimes b) \otimes c = a \otimes (c \otimes b).$$

In (S), replace a, b, c, d respectively by $a, 0, c, b$; use I and III.

V. *The set obtained by omitting* 0 *from* K *is an abelian group with respect to* $\otimes$.

Let us call this set K_0. By IV, whenever $a, b, c, a \otimes b, c \otimes b,$ and $a \otimes (c \otimes b)$ are in K_0, then $(a \otimes b) \otimes c = a \otimes (c \otimes b)$. By (M) and II, whenever a and b are in K_0, there is an x in K_0 such that $a \otimes x = b$. These two results exactly agree with postulates (1), (2) in § 1 of AGF with $\otimes$ for 0 and K_0 for K; hence the theorem is proved.

This result, together with II, III implies that for every a, b in K, $a \otimes b$ is in K, and that there is an identity 1 in K such that for every a in K, $a \otimes 1 = a$.

VI. If a, b, c, d *belong to* K, *then* $a \otimes (b \oplus c) = (a \otimes b) \oplus (a \otimes c)$.

In (S), put $d = 1$, using V.

3. **Verification of the field laws.** By I, K is an abelian group with respect to $\oplus$; by V, K without 0 is an abelian group with respect to $\otimes$; by II and III $a \otimes 0 = 0 \otimes a = 0$; and by VI, the distributive law holds. Hence a field is defined by $(K; \oplus, \otimes)$.

4. **Complete independence.** The complete independence of each of the sets of postulates

$$(A1),\ (A2),\ (M),\ (S),\ (Nf); \qquad (A1),\ (A2),\ (M),\ (S),\ (N\infty)$$

is exhibited by a table, in which, for the sake of brevity, the choice of K and the sign for the N-postulate are not indicated. These are to be chosen as follows: for (Nf) + and for $(N\infty)$ −, let K be any finite field of at least two elements; for (Nf) − and for $(N\infty)$ +, let K be any infinite field.

A1	A2	M	S	$a \oplus b$	$a \otimes b$
+	+	+	+	$a+b$	ab
+	+	+	−	$a+b$	$a+b$
+	+	−	+	$a+b$	0
+	+	−	−	$a+b$	a
+	−	+	+	0	ab
+	−	+	−	0	$a+b$
+	−	−	+	0	0
+	−	−	−	0	a

A1	A2	M	S	$a \oplus b$	$a \otimes b$
−	+	+	+	b	$a+b$
−	+	+	−	b	b
−	+	−	+	$b+1$	$a-1$
−	+	−	−	$b+1$	a
−	−	+	+	$a+1$	b
−	−	+	−	$a(b+1)$	b
−	−	−	+	$a(b+1)$	0
−	−	−	−	$a+1$	a

CORNELL UNIVERSITY, ITHACA, N. Y.,
August, 1931.

NOTE ON MULTIPLE DIRICHLET AND MULTIPLE FACTORIAL SERIES.[1]

By C. Raymond Adams.

A recent paper[2] in which the writer developed some of the fundamental properties of multiple factorial series left open the question of the validity for multiple series of the natural analogue of the well known theorem concerning simple series: *A factorial series in x and the associated Dirichlet series both converge or both diverge at each point x ($\neq 0, -1, -2, \cdots$) of the complex plane.* It turns out that the natural analogue is not valid, so that in this respect the theory of the multiple series exhibits an interesting difference from that of the simple series. The chief purpose of this note is to investigate the relation between the regions of convergence[3] of the two multiple series; we shall show that, except for boundary points, they are the same. We shall also, in large part incidentally, establish some convergence properties of multiple Dirichlet series and shall point out one or two extensions of the results of our earlier paper.

For simplicity we confine our discussion to *double* series; it need be modified only in a formal way to cover series of multiplicity $n > 2$. The basis of the proofs is Hardy's generalization to multiple series of Abel's partial summation device,[4] of which the fundamental identity for double series is

$$(1)\qquad \sum_{i=1,j=1}^{s,t} a_{ij} b_{ij} = \sum_{i=1,j=1}^{s-1,t-1} S_{ij} \Delta_{11} b_{ij} - \sum_{i=1}^{s-1} S_{it} \Delta_{10} b_{it} - \sum_{j=1}^{t-1} S_{sj} \Delta_{01} b_{sj} + S_{st} b_{st},$$

where, in general, $S_{ij} = \sum_{p=1,q=1}^{i,j} a_{ij}$ and

$$\Delta_{10} b_{ij} = b_{i+1,j} - b_{ij}, \quad \Delta_{01} b_{ij} = b_{i,j+1} - b_{ij}, \quad \Delta_{11} b_{ij} = \Delta_{10}(\Delta_{01} b_{ij}).$$

1. Convergence of double Dirichlet series. Consider the series

$$(2)\qquad \sum_{i,j=1}^{\infty} c_{ij} e^{-\lambda_i x - \mu_j y},$$

[1] Received July 20, 1931.—Presented to the American Mathematical Society, September 9, 1931.

[2] Adams, "On multiple factorial series", these *Annals*, ser. 2, vol. 32 (1931), pp. 67-82.

[3] Convergence as defined by Pringsheim, *Encyklopädie der mathematischen Wissenschaften*, vol. I, A 3, pp. 97–98.

[4] Hardy, "On the convergence of certain multiple series", *Proceedings of the London Mathematical Society*, ser. 2, vol. 1 (1903), pp. 124–128.

in which $\{\lambda_i\}$ is a sequence of real numbers, monotonically increasing without limit; $\{\mu_j\}$ is a like sequence; the c_{ij} are constants; and x and y are complex variables. The basic convergence property of this series is given by

THEOREM 1. *If the partial sum S_{ij} of a double Dirichlet series* (2) *is bounded for the place (x_0, y_0), the series is convergent with S_{ij} bounded*[5] *for every place (x_1, y_1) satisfying the conditions*[6] $R(x_1) > R(x_0)$, $R(y_1) > R(y_0)$.

We shall omit the proof of this theorem, since it follows precisely the lines of the slightly less simple proof of Theorem 2 of the next section. It is worthy of note, however, that in the same manner we can establish a similar theorem for double factorial series, which is stronger than Theorem 1 of our earlier paper.[2]

From Theorem 1 we conclude that *in general there exist two related numbers σ_1 and σ_2 such that the series* (2) *converges with S_{ij} bounded for $R(x) > \sigma_1$, $R(y) > \sigma_2$ and fails to converge with S_{ij} bounded for $R(x) < \sigma_1$, $R(y) < \sigma_2$; in particular either one or both of these numbers may become positively or negatively infinite.* We also infer that *within the related half-planes of convergence with S_{ij} bounded, each row and column of* (2) *is convergent,* so that the double series may be summed either by rows or by columns in this region.[7]

2. **Uniform convergence of double Dirichlet series in a sectorial region.** We shall now prove the following analogue of a familiar theorem on simple series.[8]

THEOREM 2. *If the partial sum S_{ij} of a double Dirichlet series* (2), *with*[9] $\lambda_1 \geqq 0$ *and* $\mu_1 \geqq 0$, *is bounded for the place (x_0, y_0), the series converges uniformly with S_{ij} uniformly bounded for (x, y) in the region D defined by the inequalities*

$$|\operatorname{am}(x - x_0)| \leqq \theta < \pi/2, \qquad R(x - x_0) \geqq \epsilon > 0;$$
$$|\operatorname{am}(y - y_0)| \leqq \theta, \qquad R(y - y_0) \geqq \epsilon.$$

Proof. We set

$$a_{ij} = c_{ij}\, e^{-\lambda_i x_0 - \mu_j y_0}, \qquad b_{ij}(x, y) = e^{\lambda_i (x_0 - x) + \mu_j (y_0 - y)},$$

---

[5] S_{ij} is used, in a generic sense, to designate the sum of the terms in the first i rows and the first j columns of any double series to which the phrase "convergent with S_{ij} bounded" is applied.

[6] $R(x)$ denotes the *real part* of x.

[7] See Pringsheim, *Vorlesungen über Zahlen- und Funktionenlehre,* Leipzig, I. 2, 1916, p. 457.

[8] See Hardy and Riesz, *The General Theory of Dirichlet's Series,* Cambridge, 1915, p. 3, Theorem 2. We shall refer to this tract as H.-R.

[9] The necessity for this restriction is evident. At the same time it should be observed that the product of the series (2) by an exponential factor e^{-lx-my}, with suitably chosen values of the constants l and m, will always satisfy this condition.

and propose to show that under the hypotheses each term on the right in (1) approaches a limit uniformly for (x, y) in D as s and t become infinite simultaneously but independently.

Since by hypothesis there exists a constant M for which the inequality

$$|S_{ij}| < M \qquad (i, j = 1, 2, 3, \cdots) \tag{3}$$

is satisfied, the first term on the right in (1) will surely approach a limit uniformly if the double series

$$\sum_{i,j=1}^{\infty} |\Delta_{11}\, b_{ij}(x, y)| \tag{4}$$

converges uniformly in D. But this series is the formal product of the two simple series[10]

$$\sum_{i=1}^{\infty} |\Delta\, e^{\lambda_i(x_0-x)}|, \qquad \sum_{j=1}^{\infty} |\Delta\, e^{\mu_j(y_0-y)}|, \tag{5}$$

and will converge uniformly in D if both of the series (5) do likewise. By virtue of a well known inequality[11] and the relation

$$|x-x_0|/R(x-x_0) \leqq \sec\theta,$$

which by hypothesis is satisfied in D, the first of series (5) is dominated by the series

$$(\sec\theta) \sum_{i=1}^{\infty} \Delta\, e^{-\lambda_i R(x-x_0)}.$$

This series converges to its sum, $-(\sec\theta)\, e^{\lambda_1 R(x_0-x)}$, uniformly in D, since the function $e^{\lambda_n R(x_0-x)}$ tends to zero uniformly in D. Consequently the first of series (5), and likewise the second, converges uniformly in D.

The remaining terms on the right in (1) all approach zero uniformly in D: the fourth on account of (3) and the relation

$$\lim_{s,t\to\infty} b_{st}(x, y) = 0,$$

which is seen to hold uniformly in D; the second because we have

$$\left| \sum_{i=1}^{s-1} S_{it}\, \Delta_{10}\, b_{it}(x, y) \right| < M\, |e^{\mu_t(y_0-y)}| \cdot \sum_{i=1}^{s-1} |\Delta\, e^{\lambda_i(x_0-x)}|,$$

whose limit, zero, is approached uniformly in D since the expression itself is the product of a function of t which tends to zero uniformly in D and

[10] Δ operates respectively on i and j.

[11] See H.-R., p. 3, Lemma 2.

a function of s which has just been proved to tend uniformly to a function bounded in D; and the third for similar reasons.

Hence the left member of (1) approaches a limit uniformly in D; that this sum is uniformly bounded in D follows at once from the above discussion. This completes the proof of our theorem.[12]

3. **The relation between the regions of convergence of double factorial and associated double Dirichlet series.** We now come to the main object of this note, which is to demonstrate the two following theorems.

THEOREM 3. *The related right half-planes of convergence with S_{ij} bounded of the double Dirichlet series*

$$\sum_{i,j=1}^{\infty} c_{ij}\, i^{-x} j^{-y} \tag{2'}$$

and of the double factorial series

$$\sum_{i,j=1}^{\infty} \frac{i!\, j!\, c_{ij}}{x(x+1)\cdots(x+i)\,y(y+1)\cdots(y+j)} \tag{6}$$

are identical.

Proof. It is sufficient to establish two facts: (a) that at each interior place[13] of the region of convergence with S_{ij} bounded of the series (2') the series (6) converges with S_{ij} bounded, and (b) that at each interior place of the region of convergence with S_{ij} bounded of the series (6) the series (2') converges with S_{ij} bounded. Detailed consideration need be given here only to (a), since (b) can be treated in a similar way.

Let us then assume (x, y) to be any particular place satisfying the hypothesis of (a) and set

$$a_{ij} = \frac{c_{ij}}{i^x j^y}, \qquad b_{ij} = \frac{i!\, j!\, i^x j^y}{x(x+1)\cdots(x+i)\,y(y+1)\cdots(y+j)}.$$

The partial sum S_{ij} of the double series $\sum_{i,j=1}^{\infty} a_{ij}$ is bounded, let us say by M, and we shall show that each term on the right in (1) approaches a limit as s and t simultaneously but independently become infinite.

---

[12] We take this opportunity to remark that, by a process amounting essentially to translation, the unnatural restriction that χ and ψ be positive can be removed from Theorem 7 of our earlier paper, loc. cit. It is of course understood that if either χ or ψ or both are negative, such of the points $0, -1, -2, \cdots$ as lie in the right half-planes defined, together with neighborhoods of them, shall be excluded from the region.

[13] By this term is meant a place (x, y) for which *both* x and y are interior points of a pair of related right half-planes of convergence. Throughout this section x and y are understood to be different from $0, -1, -2, \cdots$.

To prove that the first term approaches a limit is to prove the convergence of the double series

$$\sum_{i,j=1}^{\infty} S_{ij}\, \Delta_{11}\, b_{ij}. \tag{7}$$

That this series converges absolutely is easily seen, for S_{ij} is bounded and

$$\sum_{i,j=1}^{\infty} |\, \Delta_{11}\, b_{ij}\,|$$

may be regarded as the product of the two simple series[10]

$$\sum_{i=1}^{\infty} \left| \Delta \frac{i!\, i^x}{x(x+1)\cdots(x+i)} \right|, \quad \sum_{j=1}^{\infty} \left| \Delta \frac{j!\, j^y}{y(y+1)\cdots(y+j)} \right|, \tag{8}$$

each of which is known to converge[14].

As for the fourth term on the right in (1), we have by hypothesis that $\lim S_{st}$ exists as s and t become infinite in the specified manner, and

$$\lim b_{st} = \Gamma(x)\, \Gamma(y).$$

The second term may be written

$$\frac{t!\, t^y}{y(y+1)\cdots(y+t)} \cdot \sum_{i=1}^{s-1} S_{it}\, \Delta \frac{i!\, i^x}{x(x+1)\cdots(x+i)}.$$

Since the limit of the first factor on the right exists and equals $\Gamma(y)$, it remains to show that the limit of the second factor exists. This must surely be the case if the ordinary double limit of

$$\sum_{i=1}^{s-1} S_{it}\, \Delta \frac{i!\, i^x}{x(x+1)\cdots(x+i)}, \tag{9}$$

as s and t become infinite independently but not necessarily simultaneously, exists. Now (x, y) being an interior place of the related half-planes of convergence of (2′), each row and column of the series $\sum_{i,j=1}^{\infty} a_{ij}$ converges; whence

$$\lim_{t \to \infty} S_{it} \text{ exists} \qquad (i = 1, 2, 3, \cdots),$$

and

$$\lim_{i \to \infty} S_{it} \text{ exists} \qquad (t = 1, 2, 3, \cdots).$$

Consequently the simple limit of (9), as t becomes infinite, exists for each s. Moreover, the simple limit of (9), as s becomes infinite, exists *uniformly*

[14] See Landau, "Über die Grundlagen der Theorie der Fakultätenreihen", *Sitzungsberichte der Münchener Akademie* (math.-phys. Kl.), vol. 36 (1906), pp. 168–169.

with respect to t, since S_{it} by hypothesis is bounded and the first of series (8) converges. Hence by a familiar theorem the double limit of (9) exists.

That the third term on the right in (1) approaches a limit may be shown in the same manner; it then follows that the series (6) is convergent for the place (x, y). Moreover, the discussion shows that each term on the right in (1) is a bounded function of s and t; thus the partial sum of the series (6) is bounded for the place (x, y), and the proof of (a) is complete.

THEOREM 4. *The series* (2′) *and* (6) *need not, although they may, both converge with* S_{ij} *bounded, or fail so to converge, at each boundary place*[15] *of the related half-planes of convergence.*

To prove this theorem it is enough to exhibit examples of each possibility. A class of series illustrating the second possibility has already been remarked;[16] examples of the first may readily be constructed in the following manner.

Let us suppose that for the place (1,1) the series (2′) has the partial sum

$$S_{ij} = (-1)^j/i.$$

It is quite clear that the coefficients c_{ij} can be chosen in a unique manner to bring this about. Then $|S_{ij}|$ is at most equal to 1, and by Theorem 1 the series (2′) converges with S_{ij} bounded for all (x, y) satisfying the conditions $R(x) > 1$, $R(y) > 1$. Moreover, it is evident that for the place $(1, 1)$ the series (2′) converges to zero.

That the associated series (6) does not converge at (1,1) may be shown by proving that the first, third, and fourth terms on the right in (1) tend to limits while the second does not. As in the proof of Theorem 3 the double series (7) converges absolutely, and we have, as s and t become infinite in any manner whatsoever,

$$\lim S_{st}\, b_{st} = 0;$$

thus the first and fourth terms have limits. In the above demonstration of Theorem 3 the proof of the existence of a limit for the third term on the right in (1) makes use of only two properties of the series (2′): (i) $\lim_{i\to\infty} S_{ij}$ exists, for each j; and (ii) S_{ij} is bounded. The series (2′) of the present example possesses both of these properties; hence the third term on the right in (1) has a limit.

[15] By this term is meant a place (x, y) for which *either or both* x and y are boundary points of a pair of related right half-planes of convergence.

[16] See Adams, "Factorial series in two variables", *Bulletin of the American Mathematical Society*, vol. 34 (1928), p. 474, Theorem 6.

To prove that the second term does not tend to a limit it is sufficient, since

$$\lim_{t\to\infty} \frac{t!\, t^y}{y(y+1)\cdots(y+t)}\bigg|_{y=1} = \Gamma(1)$$

is not zero, to show that the limit of (9), as s and t simultaneously but independently become infinite, does not exist. But we have

$$\Delta \frac{i!\, i^x}{x(x+1)\cdots(x+i)}\bigg|_{x=1} = \frac{1}{(i+1)(i+2)},$$

so that the sum (9) is

$$\frac{1}{2\cdot 3}S_{1t} + \frac{1}{3\cdot 4}S_{2t} + \cdots + \frac{1}{s(s+1)}S_{s-1,t}$$
$$= (-1)^t\left[\frac{1}{1\cdot 2\cdot 3} + \frac{1}{2\cdot 3\cdot 4} + \cdots + \frac{1}{(s-1)s(s+1)}\right],$$

which oscillates between values $>\frac{1}{8}$ and values $<-\frac{1}{8}$ and therefore has no limit.

It is manifestly impossible to draw any more general conclusion concerning the question at issue for a place (x, y) of which one of the points, say x, lies on the boundary of a half-plane of convergence, while y is an interior point of the related half-plane of convergence, for it would frequently happen, although it might not, that both x and y were interior points of a different pair of related half-planes of convergence. Nevertheless it seems worth while to make the following observation, the truth of which is apparent from the above discussion: *if at a place* (x, y) *for which* x *or* y *or both lie on the boundary of a half-plane of convergence, either series* (2′) *or series* (6) *converges with* S_{ij} *bounded and each of its rows and columns converges, then the other series has these same properties.*

Brown University, Providence, R. I.,
July 17, 1931.

Note inserted in proof: There has just come to our attention the article by Leja, "Sur les séries de Dirichlet doubles", *Comptes-Rendus du 1 Congrès des Mathématiciens des Pays Slaves*, Warsaw, 1930, pp. 140–158. The first part of Leja's Theorem 1 is identical with our Theorem 1; the second part, which states that the convergence is uniform in the neighborhood of each place (x_1, y_1) satisfying the conditions $R(x_1) > R(x_0)$, $R(y_1) > R(y_0)$, is a consequence of our Theorem 2.

ON DEFERRED CESÀRO MEANS.[1]

By Ralph Palmer Agnew.[2]

1. In recent papers[3] the writer has found that nearly all of the transformations used in the theory of summability have undesirable features. For example the Cesàro transformation of any given positive order increases ultimate bounds and oscillations of certain sequences of functions, and does not always preserve uniform convergence, or continuous convergence, or mean-square convergence, of sequences of functions. It is the object of this paper to consider modified Cesàro transformations which have useful properties not possessed by Cesàro's and other well known transformations.

2. The transformations which we shall consider are of the type

$$\sigma_n = \sum_{k=1}^{\infty} a_{nk} s_k \tag{2.1}$$

which assign to a given sequence $\{s_n\}$ the value $\lim_{n\to\infty} \sigma_n$ when this limit exists. With reference to (2.1), the following conditions are of importance:

$$\sum_{k=1}^{\infty} |a_{nk}| \text{ is bounded for all } n, \tag{2.2}$$

$$\text{for each } k, \lim_{n\to\infty} a_{nk} = 0, \tag{2.3}$$

$$\lim_{n\to\infty} \sum_{k=1}^{\infty} a_{nk} = 1, \tag{2.4}$$

$$a_{nk} \geqq 0 \text{ for all } n \text{ and } k, \tag{2.5}$$

$$\text{for each } k, \ a_{nk} = 0 \text{ for almost all } n,[4] \tag{2.6}$$

$$\sum_{k=1}^{\infty} a_{nk} = 1 \text{ for all } n. \tag{2.7}$$

The first three of these conditions are, by the Silverman-Toeplitz theorem, necessary and sufficient for *regularity*[5] of (2.1); the last three are sufficient

[1] Received May 20, 1931. Presented to the American Mathematical Society, September 11, 1931.

[2] National Research Fellow.

[3] Transactions of the American Mathematical Society, vol. 32 (1930), pp. 669–708; ibid., vol. 33 (1931), pp. 411–424; American Journal of Mathematics, vol. 53 (1931), pp. 204–216.

[4] I. e. for each k, $a_{nk} = 0$ except for a finite number, depending on k, of values of n.

[5] For definitions of *regularity, equivalence, mutual-consistency,* etc., see W. A. Hurwitz, Bulletin of the American Mathematical Society, vol. 28 (1922), pp. 17–36.

to ensure (2.1) having useful properties in addition to regularity. The Cesàro arithmetic mean transformation

$$(M) \qquad M_n = (s_1 + s_2 + \cdots + s_n)/n$$

satisfies all of these conditions except (2.6); it is the failure of (M) to satisfy (2.6) which accounts for its having the undesirable features mentioned in § 1.

3. While it is the primary object of this paper to discuss modifications of (M) which satisfy (2.6), we shall consider general transformations of the form

$$(D) \qquad D_n = \frac{s_{p_n+1} + s_{p_n+2} + \cdots + s_{q_n}}{q_n - p_n}$$

where $\{p_n\}$ and $\{q_n\}$ are sequences of non-negative integers satisfying the conditions

$$(3.1) \qquad p_n < q_n, \qquad n = 1, 2, 3, \cdots$$

$$(3.2) \qquad \lim_{n\to\infty} q_n = +\infty.$$

Such a transformation (D) will be called the *deferred Cesàro Transformation* determined by $\{p_n\}$ and $\{q_n\}$ and will, when a completely descriptive notation is required, be denoted by $D(\{p_n\}, \{q_n\})$. We see that (M) involves means of *first* elements of $\{s_n\}$; and that (D) involves, except in case $p_n = 0$ all n, means of *deferred* elements of $\{s_n\}$. It may be noted that $D(\{n-1\}, \{n\})$ is the identity, and that $D(\{0\}, \{n\})$ is (M) itself.[6]

For (D) we have, in the notation of (2.1), $a_{nk} = 1/(q_n - p_n)$ if $p_n < k \leqq q_n$, and $a_{nk} = 0$ if $k \leqq p_n$ or $k > q_n$. Since each (D) satisfies (2.2), (2.3), and (2.4), *every* (D) *is regular*. We see also that *each* (D) *satisfies* (2.5) *and* (2.7), *and in order that* (D) *may satisfy* (2.6), *it is necessary and sufficient that* $\lim_{n\to\infty} p_n = +\infty$.

We shall say that (D) is *properly deferred* when $\{p_n\}$ and $\{q_n\}$ satisfy, in addition to (3.1) and (3.2), the condition

$$(3.3) \qquad p_n/(q_n - p_n) \text{ is bounded for all } n;$$

such a transformation will be called a *proper* (D).

4. An important property of the set of proper (D)'s is exhibited by the following theorem.

[6] Mollerup, Det Kgl. Danske Videnskabernes Selskab. Math.-fys. Medd, Vol. 3 (1920), No. 8, defines and discusses summability $M(r)$ which involves means of the form $D(\{n\}, \{q_n\})$. Mollerup's results partially include Theorem 4.1 of the present paper.

THEOREM 4.1. *In order that* (D) *may include* (M), *it is necessary and sufficient that* (D) *be proper.*

Let $\{s_n\}$ be a given sequence; then for any transformation (D) we have

$$D_n = (s_{p_n+1} + s_{p_n+2} + \cdots + s_{q_n})/(q_n - p_n)$$
$$= -[p_n/(q_n - p_n)]\,[(s_1 + \cdots + s_{p_n})/p_n] + [q_n/(q_n - p_n)]\,[(s_1 + \cdots + s_{q_n})/q_n]$$

so that

$$D_n = -[p_n/(q_n - p_n)]\,M_{p_n} + [q_n/(q_n - p_n)]\,M_{q_n}. \tag{4.2}$$

Regarding (4.2) as a transformation of the form (2.1) which carries $\{M_n\}$ into $\{D_n\}$, we see that (4.2) satisfies (2.3) and (2.4); and that (4.2) satisfies (2.2) when and only when $[p_n/(q_n - p_n) + q_n/(q_n - p_n)]$ is bounded, i. e. when and only when $p_n/(q_n - p_n)$ is bounded and (D) is proper. Thus (4.2) is regular when and only when (D) is proper. Since (M) has an inverse, (D) includes (M) when and only when (4.2) is regular, and the theorem is proved.

From Theorem 4.1 it follows that *any proper* (D) *and* (M) *are mutually consistent.*

5. However it is not true that any two proper (D)'s are mutually consistent. In fact the sequence $\{x_n\}$, whose elements $1, -1, 2, -2, 3, -3, 4, \cdots$ are the partial sums of the series $1-2+3-4+5-6+\cdots$, is assigned the value 0 by $D(\{2n\}, \{4n\})$ and the value $\frac{1}{2}$ by $D(\{2n-1\}, \{4n-1\})$. Since $\{x_n\}$ is assigned the value $\frac{1}{4}$ by the Cesàro mean (C_2) of order 2, this example shows that a proper (D) and (C_2) are not necessarily mutually consistent.

When a method is not consistent with (C_2), it is not consistent with any method which includes (C_2); hence certain proper (D)'s are open to the serious objection that they are not consistent with the Euler-Abel-Poisson power series method of analytic continuation.

These objections do not apply to (D)'s which are included by (M), for all such (D)'s must be consistent with each other and with all methods which are consistent with (M).

We will now show that (M) includes certain sets of transformations of the form (D).

6. Transformations which have triangular matrices with inverses, i. e transformations of the form (2.1) for which $a_{nk} = 0$ when $k > n$ and $a_{nn} \neq 0$, play an extremely important role in the theory of summability. Hence the following theorems involving transformations of the form $D(\{p_n\}, \{n\})$ are of especial interest.

THEOREM 6.1. *The transformation* (M) *includes every transformation of the form* $D(\{p_n\}, \{n\})$.

Let $\{s_n\}$ be summable $D(\{p_n\}, \{n\})$ to L so that $\lim D_n = L$; we shall show that $\lim M_n = L$. Letting, for each n, $p_n = n^{(1)}$, $p_{n^{(1)}} = n^{(2)}$, $p_{n^{(2)}} = n^{(3)}, \cdots$, we may write

$$\begin{aligned} M_n &= \frac{1}{n}(s_1 + \cdots + s_{n^{(1)}} + s_{n^{(1)}+1} + \cdots + s_n) \\ &= \frac{s_1 + s_2 + \cdots + s_{n^{(1)}}}{n} + \frac{n - n^{(1)}}{n} D_n \\ &= \frac{s_1 + \cdots + s_{n^{(2)}}}{n} + \frac{n^{(1)} - n^{(2)}}{n} D_{n^{(1)}} + \frac{n - n^{(1)}}{n} D_n. \end{aligned}$$

This reduction may be continued until, for some positive integer h depending on n, we obtain

$$\begin{aligned} M_n = \frac{n^{(h)} - n^{(h+1)}}{n} D_{n^{(h)}} + \frac{n^{(h-1)} - n^{(h)}}{n} D_{n^{(h-1)}} + \cdots \\ \cdots + \frac{n^{(1)} - n^{(2)}}{n} D_{n^{(1)}} + \frac{n - n^{(1)}}{n} D_n \end{aligned}$$

where $n^{(h)} \geqq 1$ and $n^{(h+1)} = 0$. The preceding relation, which holds for each n, may be regarded as a transformation of the form (2.1) which carries $\{D_n\}$ into $\{M_n\}$. Since $n^{(r)} > n^{(r+1)}$, $r = 1, 2, \cdots, h$, and $n^{(h+1)} = 0$, this transformation evidently satisfies (2.2) and (2.4); and for a fixed k, the coefficient of D_k is either zero or a fraction of which the denominator is n and the numerator is $\leqq k$ so that (2.3) holds. Therefore, by the Silverman-Toeplitz theorem, $\lim M_n = L$ follows from $\lim D_n = L$ and Theorem 6.1 is proved.

Combining Theorems 4.1 and 6.1, we obtain

THEOREM 6.2. *A transformation of the form $D(\{p_n\}, \{n\})$ is equivalent to (M) when and only when $p_n/(n - p_n)$ is bounded for all n.*

As a corollary of Theorem 6.2 we obtain the following result: *If θ is a constant, $0 \leqq \theta < 1$, and $[\theta n]$ is the greatest integer $\leqq \theta n$, then the transformation*

$$D_n^{(\theta)} = \frac{s_{[\theta n]+1} + s_{[\theta n]+2} + \cdots + s_n}{n - [\theta n]} \tag{6.3}$$

is equivalent to (M). With the matrix of (6.3) are associated two lines, one depending on θ, which bound a closed sector containing all of the non-vanishing elements of the matrix; by taking θ sufficiently close to 1, we can make the angle at the vertex of this closed sector as small as we please.

7. Two transformations $\sigma_n = \sum a_{nk} s_k$ and $\sigma_n' = \sum a_{nk}' s_k$ in which a_{nk} and a_{nk}' differ for at most a finite set of values of n are equivalent. Hence we obtain, from Theorem 6.1, the following lemma.

LEMMA 7.1. *The transformation* (M) *includes every transformation* $D(\{p_n\}, \{q_n\})$ *for which* $q_n = n$ *for almost all* n.

THEOREM 7.2. *The transformation* (M) *includes every transformation* $D(\{p_n\}, \{q_n\})$ *for which* $\{q_n\}$ *contains almost all positive integers.*

Let $\{s_n\}$ be summable $D(\{p_n\}, \{q_n\})$ to L so that $\lim D_n = L$, and choose an integer K so great that $\{q_n\}$ contains all integers greater than K. Then we may set $i_1 = i_2 = \cdots = i_K = 1$ and determine for each $n > K$ an index i_n such that $q_{i_n} = n$. Since $\lim_{n\to\infty} i_n = +\infty$, it follows from $\lim D_n = L$ that $\lim D_{i_n} = L$; hence $\{s_n\}$ is summable $D(\{p_{i_n}\}, \{q_{i_n}\})$ to L, and the theorem follows from Lemma 7.1.

THEOREM 7.3. *The transformation* (M) *fails to contain* $D(\{p_n\}, \{q_n\})$ *if there exists an increasing sequence* $\{\alpha_n\}$ *of integers whose elements belong to neither* $\{p_n\}$ *nor* $\{q_n\}$.

Let $\bar{M}_n = 0$, $n \neq \alpha_1, \alpha_2, \cdots$, and $\bar{M}_{\alpha_n} = x_n$ where $\{x_n\}$ is a divergent sequence; and let $\{\bar{s}_n\}$ be the sequence which is carried by (M) into $\{\bar{M}_n\}$. Using (4.2) and the fact that p_n and q_n always differ from $\alpha_1, \alpha_2, \cdots$, we see that $D(\{p_n\}, \{q_n\})$ evaluates $\{\bar{s}_n\}$ to 0. Since (M) fails to evaluate $\{\bar{s}_n\}$, the theorem is proved.

8. In § 4 we gave a simple necessary and sufficient condition that (D) may include (M). In § 7 we have a simple sufficient condition and a simple necessary condition that (M) may include (D). It would be desirable to have simple necessary and sufficient conditions that (M) may include (D), but the writer has been able to obtain such conditions only for (D)'s of restricted forms. The theorems and examples below give criteria which are easily applied and throw some light on the general problem.

THEOREM 8.1. *In order that* (M) *may include* $D(\{p_n\}, \{q_n\})$ *where* $\{p_n\}$ *is for almost all* n *equal to a given non-negative integer* p, *it is necessary and sufficient that* $\{q_n\}$ *contain almost all positive integers.*

Sufficiency follows from Theorem 7.2 and necessity from Theorem 7.3.

THEOREM 8.2. *In order that* (M) *may include* $D(\{n-1\}, \{q_n\})$ *where* $q_n - n$ *increases monotonely with* n, *it is necessary and sufficient that* $q_n - n$ *be bounded.*

To establish sufficiency, we note that $q_n - n$ must have a limit, say L, and that $q_n - n = L$ for almost all n; hence $\{q_n\}$ contains almost all positive integers and we may apply Theorem 7.2.

To establish necessity, suppose $q_n - n$ increases monotonely with n and is not bounded; we will show that (M) fails to include (D). Let $n_1 = 1$; let n_2 be the smallest integer n for which $q_n - n > q_{n_1} - n_1$; let n_3 be the smallest integer n for which $q_n - n > q_{n_2} - n_2$; and in general let $n_{\alpha+1}$ be the smallest integer n for which $q_n - n > q_{n_\alpha} - n_\alpha$. Then $1 = n_1 < n_2$

$< n_3 < \cdots$. We find that we can set $\bar{s}(q_{n_i}) = q_{n_i}$, $i = 1, 2, 3, \cdots$ and choose the remaining elements of a sequence $\{\bar{s}(n)\} \equiv \{\bar{s}_n\}$ so that the sequence $\{\bar{s}_n\}$ is carried by (D) into a sequence consisting exclusively of 1's, i. e. $\bar{D}_n = 1$ for all n. Thus $\{\bar{s}_n\}$ is summable (D) to 1. But $\{\bar{s}_n\}$ cannot be summable (M) since the condition $\lim (\bar{s}_n/n) = 0$ is not satisfied. Therefore (M) does not include (D) and necessity is established.

It follows from Theorems 4.1 and 8.2 that no $D(\{n-1\}, \{q_n\})$ for which $q_n - n$ is monotone increasing can be equivalent to (M). For example, each of the transformations $D' \equiv D(\{n-1\}, \{2n-1\})$ and $D'' \equiv D(\{n-1\}, \{2n\})$ includes (M), but (M) includes neither D' nor D''. It is easy to show that D' and D'' are not mutually consistent; in fact, corresponding to any two complex numbers α and β, we can find a sequence which D' evaluates to α and D'' evaluates to β.

THEOREM 8.3. *In order that* (M) *may include* $D(\{n-1\}, \{q_n\})$ *where* $q_n = n$, $n \neq h_1, h_2, \cdots$ *and* $q_{h_i} = h_{i+1} - 1$, $i = 1, 2, \cdots, \{h_n\}$ *being an increasing sequence of integers for which* $h_n > n$, *it is necessary and sufficient that* q_n/n *be bounded for all* n.

Corresponding to each index n, let $i = i(n)$ be such that $h_i \leqq n < h_{i+1}$. Then, given a sequence $\{s_n\}$, we may write[7]

$$M_n = \frac{1}{n}[s_1 + \cdots + s_{h_1-1} + (s_{h_1} + \cdots + s_{h_2-1}) + \cdots + (s_{h_i} + \cdots + s_{h_{i+1}-1}) - (s_{n+1} + s_{n+2} + \cdots + s_{h_{i+1}-1})].$$

Using the definition of (D), we obtain

$$M_n = \frac{1}{n}(D_1 + \cdots + D_{h_1-1}) + \frac{h_2 - h_1}{n} D_{h_1} + \frac{h_3 - h_2}{n} D_{h_2} + \cdots + \frac{h_{i+1} - h_i}{n} D_{h_i} - \frac{1}{n}(D_{n+1} + D_{n+2} + \cdots + D_{h_{i+1}-1}).$$

The preceding relation holds for each n and defines a transformation of the form (2.1) which carries $\{D_n\}$ into $\{M_n\}$. This transformation satisfies (2.3) and (2.4); and satisfies (2.2) when and only when $(2h_{i+1} - n - 2)/n$ is bounded for all n, i. e. when and only when (h_{i+1}/h_i) is bounded for all n and therefore when and only when q_n/n is bounded for all n. Since D has an inverse, Theorem 8.3 follows from an application of the Silverman-Toeplitz theorem.

THEOREM 8.4. *The transformation* (M) *includes* $D(\{n-1\}, \{q_n\})$ *where* $q_n = n$, $n \neq h_1, h_2, \cdots$, *and* $q_{h_i} = h_{i+1}$, $i = 1, 2, 3, \cdots, \{h_n\}$ *being any increasing sequence of integers for which* $h_n > n$.

[7] In case $n = h_{i+1} - 1$, the terms with subscripts $n+1, n+2, \cdots, h_{i+1}-1$ do not appear.

Corresponding to each index n, let $i = i(n)$ be such that $h_i \leqq n < h_{i+1}$. Then we may, given a sequence $\{s_n\}$, write[8]

$$M_n = \frac{1}{n}\,[s_n + s_{n-1} + \cdots + s_{h_i+1} + (s_{h_i} + \cdots + s_{h_{i-1}}) + s_{h_{i-1}-1} + \cdots + s_{h_{i-2}+1} + (s_{h_{i-2}} + \cdots + s_{h_{i-3}}) + \cdots + (s_{h_{\delta+1}} + \cdots + s_{h_\delta}) + K_n]$$

where δ is 2 or 1 according as i is odd or even, and $K_n = s_1 + s_2 + \cdots + s_{h_\delta - 1}$. Setting $T_n = M_n - K_n/n$, and using the definition of (D), we obtain

$$T_n = \frac{1}{n}\,(D_n + D_{n-1} + \cdots + D_{h_i+1}) + \frac{h_i - h_{i-1} + 1}{n}\,D_{h_{i-1}} + \frac{1}{n}\,(D_{h_{i-1}-1} + \cdots + D_{h_{i-2}+1}) + \frac{h_{i-2} - h_{i-3} + 1}{n}\,D_{h_{i-3}} + \cdots + \frac{h_{\delta+1} - h_\delta + 1}{n}\,D_{h_\delta}.$$

The preceding relation defines a regular transformation from $\{D_n\}$ to $\{T_n\}$; hence if $D_n \to L$, then $T_n \to L$ and, since K_n is bounded, $M_n \to L$. Therefore (M) includes (D) and the theorem is proved.

The preceding result shows that the condition

(8.5) $\qquad q_n/n$ is bounded for all n

is not necessary in order that (M) may include $D(\{n-1\}, \{q_n\})$. That (8.5) is not sufficient in order that (M) may include $D(\{n-1\}, \{q_n\})$ follows from Theorem 8.2 since we can obviously define a sequence $\{q_n\}$ satisfying (8.5) for which $q_n - n$ becomes infinite monotonely.

9. We now turn our attention to transformations of the form $D(\{n-1\}, \{q_n\})$ where $\{q_n\}$ is monotone increasing. In case $q_n - n$ is monotone increasing, Theorem 8.2 enables us to determine whether $(M) \supset (D)$; hence we are interested in the case where $\{q_n\}$ increases monotonely and, for a certain set of values of n, $q_n = q_{n+1}$.

Corresponding to a given monotone increasing sequence $\{q_n\}$, we may determine two *characteristic sequences* $\{\alpha_n\}$ and $\{\beta_n\}$ as follows: Let $\alpha_0 = \beta_0 = 0$; let $\beta_1 = q_1$ and let α_1 be the greatest n for which $q_n = \beta_1$; let $\beta_2 = q_{\alpha_1+1}$, and let α_2 be the greatest n for which $q_n = \beta_2$; and in general when α_i and β_i are determined, let $\beta_{i+1} = q_{\alpha_i+1}$ and let α_{i+1} be the greatest n for which $q_n = \beta_{i+1}$. We see that $\alpha_1 < \alpha_2 < \alpha_3 < \cdots$ and $\beta_1 < \beta_2 < \beta_3 < \cdots$ and, since the definition of $D(\{n-1\}, \{q_n\})$ requires that $q_n \geqq n$, that $\alpha_n \leqq \beta_n$, $n = 1, 2, 3, \cdots$.

[8] In case $n = h_i$, the terms with subscripts $n, n-1, \cdots, h_i+1$ do not appear.

THEOREM 9.1. *In order that* (M) *may include a* $D(\{n-1\}, \{q_n\})$ *for which* $\{q_n\}$ *is monotone increasing and the characteristic sequences satisfy the condition*

$$\alpha_1 \leqq \beta_1 < \alpha_2 \leqq \beta_2 < \alpha_3 \leqq \cdots, \tag{9.2}$$

it is necessary and sufficient that (q_n/n) *be bounded for all* n.

For each index n, choose $i = i(n)$ such that $\alpha_{i-1} \leqq n < \alpha_i$. Then for any sequence $\{s_n\}$ we may write

$$M_n = \frac{s_1 + \cdots + s_{\beta_1}}{n} + \frac{s_{\beta_1+1} + \cdots + s_{\beta_2}}{n} + \cdots$$
$$\cdots + \frac{s_{\beta_{i-1}} + \cdots + s_{\beta_i}}{n} - \frac{s_{n+1} + \cdots + s_{\beta_i}}{n}$$

and use the definition of (D) to obtain

$$M_n = \frac{\beta_1}{n} D_1 + \frac{\beta_2 - \beta_1}{n} D_{\beta_1+1} + \cdots + \frac{\beta_i - \beta_{i-1}}{n} D_{\beta_{i-1}+1} - \frac{\beta_i - n}{n} D_{n+1}.$$

The preceding transformation from $\{D_n\}$ to $\{M_n\}$ obviously satisfies (2.3) and (2.4). We find further that $\sum_{k=1}^{\infty} |a_{nk}| = 1$ or $(2\beta_i - n)/n$ according as $n = \beta_{i-1}$ or $n \neq \beta_{i-1}$; and properties of the characteristic sequences enable us to show that (2.2) holds when and only when q_n/n is bounded for all n. Hence the transformation is regular when and only when q_n/n is bounded for all n.

We shall complete the proof of Theorem 9.1 by showing that each (D) satisfying the hypotheses of Theorem 9.1 has an inverse. To show that (D) has an inverse, it is sufficient to show that the system of equations

$$s_n + s_{n+1} + \cdots + s_{q_n} = E_n, \qquad n = 1, 2, 3, \cdots \tag{9.3}$$

has, corresponding to each given sequence $\{E_n\}$ of constants, a unique solution for the elements of $\{s_n\}$. For each $n \neq \alpha_1, \alpha_2, \cdots$ we may choose $i = i(n)$ such that $\alpha_{i-1} < n < \alpha_i$. Then properties of the characteristic sequences of $\{q_n\}$ enable us to write

$$s_n + s_{n+1} + \cdots + s_{\beta_i} = E_n$$

and

$$s_{n+1} + \cdots + s_{\beta_i} = E_{n+1}$$

so that

$$s_n = E_n - E_{n+1}, \qquad n \neq \alpha_1, \alpha_2, \alpha_3, \cdots. \tag{9.4}$$

But for each index r

$$s_{\alpha_r} + s_{\alpha_r+1} + \cdots + s_{\beta_r} = E_{\alpha_r}. \tag{9.5}$$

Since $\beta_r < \alpha_{r+1}$, each, except the first, of the terms in the left member of (9.5) is determined, in terms of elements of E_n, by (9.4); hence s_{α_r} is, for each r, uniquely determined in terms of the elements of $\{E_n\}$ and existence of the inverse of (D) is established.

10. Theorems 4.1 and 9.1 enable us to establish existence of a continuum of transformations of the form $D(\{n-1\}, \{q_n\})$, with $\{q_n\}$ monotone increasing, which are equivalent to (M).

THEOREM 10.1. *Let $\lambda > 1$ be a real constant, let Λ_r be the greatest integer $\leqq \lambda^r$, and choose an index $R > 1$ such that $\Lambda_{r+1} \geqq \Lambda_r + 1$ when $r \geqq R$. If $q_n^{(\lambda)} = \Lambda_{R+1}$ when $n \leqq \Lambda_R + 1$, and, for each $r > R+1$, $q_n^{(\lambda)} = \Lambda_r$ when $\Lambda_{r-2} + 1 < n \leqq \Lambda_{r-1} + 1$, then $D(\{n-1\}, \{q_n^{(\lambda)}\})$ is equivalent to (M).*

We see that $\{q_n^{(\lambda)}\}$ is monotone increasing, and that the characteristic sequences of $\{q_n^{(\lambda)}\}$ are given by $\alpha_i = \Lambda_{R+i-1} + 1$, $\beta_i = \Lambda_{R+i}$ so that (9.2) holds. We can show that $(q_n^{(\lambda)}/n)$ is bounded for all n; hence it follows from Theorem 9.1 that $M \supset D$. On the other hand, we can show that $(n-1)/(q_n - n + 1)$ is bounded for all n; hence it follows from Theorem 4.1 that $D \supset M$ and Theorem 10.1 is proved.

If g and G are constants for which $1 < g < G$ and $\{\lambda_n\}$ is a sequence for which $g \leqq \lambda_n \leqq G$ for all n, then we may define Λ_r to be the greatest integer less than or equal to the product $\lambda_1, \lambda_2, \cdots, \lambda_r$, and proceed as in the statement and proof of Theorem 10.1 to obtain further transformations of the form $D(\{n-1\}, \{q_n\})$ which are equivalent to (M).

If λ is a positive integer, the statement of Theorem 10.1 may be simplified. We find that if λ is an integer > 1, and $\{q_n\}$ is defined by $q_n = \lambda^2$ when $n \leqq \lambda + 1$ and, when $r > 2$, $q_n = \lambda^r$ when $\lambda^{r-2} + 1 < n \leqq \lambda^{r-1} + 1$, then $D(\{n-1\}, \{q_n\})$ is equivalent to (M).

11. In §§ 8–10 we considered transformations of the form $D(\{n-1\}, \{q_n\})$. It is unnecessary to make an independent study of transformations of the form $D(\{n+h\}, \{q_n\})$ where h is 0 or some other integer $\neq -1$. Since $D(\{n-1\}, \{q_n\})$ and $D(\{n+h\}, \{q'_n\})$ are obviously equivalent when q_n and q'_n satisfy the relations $q'_n = q_{n+h+1}$ for all sufficiently great n, any question relating to a transformation of the form $D(\{n+h\}, \{q_n\})$ can be reduced at once to a question relating to a transformation of the form $D(\{n-1\}, \{q_n\})$.

PRINCETON UNIVERSITY, PRINCETON, N. J.

EXTENSION OF THE NOTION OF THE LIMIT OF THE SUM OF TERMS OF AN INFINITE SERIES.*

By G. F. Woronoi.[1]

The notion of the limit of the sum of terms of an infinite series

$$u_0+u_1+\cdots+u_n+\cdots \tag{1}$$

of real or complex numbers, can be defined, under certain conditions, even when the series (1) diverges.

The divergence of series (1) might be of various nature. We designate a series as arithmetically divergent if there exists an exponent λ such that the ratio $s_n n^{-\lambda}$ remains bounded as $n\to\infty$, where s_n is the sum of the first n terms of the series (1).

The symbol $\sum_{n=0}^{\infty} u_n$ can be interpreted as the limit, in case it exists, of the sequence $\sigma_0, \sigma_1, \cdots, \sigma_n, \cdots$ defined by

$$\sigma_n = \frac{1}{\varphi(n)}\{\varphi(n)u_0+\varphi(n-1)u_1+\cdots+\varphi(n-k)u_k+\cdots+\varphi(0)u_n\},$$

where the function $\varphi(n)$ is assumed to satisfy the conditions (i) $\varphi(n)>0$, $n=1, 2, 3, \cdots$, and (ii) the series $\varphi(0)+\varphi(1)+\cdots$ is arithmetically divergent.

The possibility of such an interpretation of the symbol $\sum_{n=0}^{\infty} u_n$ is based upon the fact that $\sum_{n=0}^{\infty} u_n = \lim_{x\to 0} f(x)$, where

$$f(x) = \sum_{n=0}^{\infty} u_n e^{-nx}, \quad x>0.$$

* Received February 15, and March 9, 1932.

[1] This is a complete translation of a note which undoubtedly will be of interest to those who work in the theory of summability. We give the translation because the note appeared in Russian, in an extremely rare publication which in all likelihood does not exist in this country [Proceedings (Dnevnik) of the XI Congress (1901) of Russian naturalists and physicians, S.-Petersburg, 1902, pp. 60–61]. G. F. Woronoi [b. 1868, d. 1908] is one of the most remarkable representatives of the Russian mathematical school. We should mention particularly his profound investigations in the asymptotic number theory [Sur une fonction transcendante et ses applications à la sommation de quelques séries, Ann. Ec. Norm. Sup., (3) **21** (1904), pp. 207–267, 459–533] and the theory of n-ary quadratic forms [Nouvelles applications des paramètres continus à la théorie des formes quadratiques, Crelle's Jour., **133** (1908), pp. 97–178, **134** (1908), pp. 198–287, **136** (1909), pp. 67–181]. J. Tamarkin.

Thus it turns out that the value of $\sum_{n=0}^{\infty} u_n$ does not depend on the choice of the function $\varphi(n)$, provided it satisfies the conditions listed above.

In an analogous fashion it is possible to interpret the symbol $\int_0^\infty u(t)\,dt$ when this integral does not exist in the usual sense. Such an extension of the notion of the integral is given by

$$\int_0^\infty u(t)\,dt = \lim_{x\to\infty} \frac{1}{\varphi(x)} \int_0^x \varphi(x-t)\,u(t)\,dt$$

and is made possible on the basis of the property

$$\int_0^\infty u(t)\,dt = \lim_{x\to 0} \left[\int_0^\infty e^{-xt}\,u(t)\,dt\right].$$

This relation is a generalization of a known theorem of Dirichlet, which states that this equality is true when the integral $\int_0^\infty u(t)\,dt$ exists in the usual sense.

Remarks of the Translator

1. If we write $\varphi(n) = P_n$, $p_0 = P_0$, $p_n = P_n - P_{n-1}$, $n = 1, 2, 3, \cdots$, we see at once that

$$\sigma_n = \frac{1}{P_n}[p_0 s_n + p_1 s_{n-1} + \cdots + p_n s_0]. \tag{2}$$

Hence Woronoi's definition of summability is identical to that used 17 years later by Nörlund,[2] and therefore includes as special cases those of Cesàro (C_α) and Riesz $R(n, \alpha)$. If we use the modern terminology we might restate the result above of Woronoi by saying that all his definitions of summability are *consistent*. The idea of the proof sketched by Woronoi (via Abel's theorem, as explained below) is precisely the same used for the proof of consistency by Zygmund[3] and Silverman-Tamarkin.[4] By Abel's theorem we mean the following statement:

---

[2] Sur une application des fonctions permutables, Lunds Univ. Årsskr., Avd. 2, 16 (1919), no. 3. Although various special cases of this definition have been used earlier, Nörlund's paper is, to the best of our knowledge, the first after that of Woronoi where the definition was stated in its complete generality. We use the term "Nörlund's transformations", although "Woronoi's transformations" would be more appropriate.

[3] O teorji srednich arythmetycznych, Mathesis Polska, 1 (1926), pp. 75–85, 119–129. Cf. also a correction, ibidem 5 (1930), p. 46.

[4] On the generalization of Abel's theorem for certain definitions of summability, Math. Zeit., 29 (1928), pp. 161–170, esp. pp. 169–170. The authors had no knowledge of the papers of Woronoi and Zygmund.

EXTENSION OF THE NOTION OF THE LIMIT OF THE SUM OF TERMS OF AN INFINITE SERIES.*

By G. F. Woronoi.[1]

The notion of the limit of the sum of terms of an infinite series

$$u_0 + u_1 + \cdots + u_n + \cdots \tag{1}$$

of real or complex numbers, can be defined, under certain conditions, even when the series (1) diverges.

The divergence of series (1) might be of various nature. We designate a series as arithmetically divergent if there exists an exponent λ such that the ratio $s_n n^{-\lambda}$ remains bounded as $n \to \infty$, where s_n is the sum of the first n terms of the series (1).

The symbol $\sum_{n=0}^{\infty} u_n$ can be interpreted as the limit, in case it exists, of the sequence $\sigma_0, \sigma_1, \cdots, \sigma_n, \cdots$ defined by

$$\sigma_n = \frac{1}{\varphi(n)} \{\varphi(n) u_0 + \varphi(n-1) u_1 + \cdots + \varphi(n-k) u_k + \cdots + \varphi(0) u_n\},$$

where the function $\varphi(n)$ is assumed to satisfy the conditions (i) $\varphi(n) > 0$, $n = 1, 2, 3, \cdots$, and (ii) the series $\varphi(0) + \varphi(1) + \cdots$ is arithmetically divergent.

The possibility of such an interpretation of the symbol $\sum_{n=0}^{\infty} u_n$ is based upon the fact that $\sum_{n=0}^{\infty} u_n = \lim_{x \to 0} f(x)$, where

$$f(x) = \sum_{n=0}^{\infty} u_n e^{-nx}, \quad x > 0.$$

* Received February 15, and March 9, 1932.

[1] This is a complete translation of a note which undoubtedly will be of interest to those who work in the theory of summability. We give the translation because the note appeared in Russian, in an extremely rare publication which in all likelihood does not exist in this country [Proceedings (Dnevnik) of the XI Congress (1901) of Russian naturalists and physicians, S.-Petersburg, 1902, pp. 60–61]. G. F. Woronoi [b. 1868, d. 1908] is one of the most remarkable representatives of the Russian mathematical school. We should mention particularly his profound investigations in the asymptotic number theory [Sur une fonction transcendante et ses applications à la sommation de quelques séries, Ann. Ec. Norm. Sup., (3) **21** (1904), pp. 207–267, 459–533] and the theory of n-ary quadratic forms [Nouvelles applications des paramètres continus à la théorie des formes quadratiques, Crelle's Jour., **133** (1908), pp. 97–178, **134** (1908), pp. 198–287, **136** (1909), pp. 67–181]. J. Tamarkin.

Thus it turns out that the value of $\sum_{n=0}^{\infty} u_n$ does not depend on the choice of the function $\varphi(n)$, provided it satisfies the conditions listed above.

In an analogous fashion it is possible to interpret the symbol $\int_0^\infty u(t)\,dt$ when this integral does not exist in the usual sense. Such an extension of the notion of the integral is given by

$$\int_0^\infty u(t)\,dt = \lim_{x\to\infty} \frac{1}{\varphi(x)} \int_0^x \varphi(x-t)\,u(t)\,dt$$

and is made possible on the basis of the property

$$\int_0^\infty u(t)\,dt = \lim_{x\to 0} \left[\int_0^\infty e^{-xt}\,u(t)\,dt\right].$$

This relation is a generalization of a known theorem of Dirichlet, which states that this equality is true when the integral $\int_0^\infty u(t)\,dt$ exists in the usual sense.

Remarks of the Translator

1. If we write $\varphi(n) = P_n$, $p_0 = P_0$, $p_n = P_n - P_{n-1}$, $n = 1, 2, 3, \cdots$, we see at once that

$$\sigma_n = \frac{1}{P_n}[p_0 s_n + p_1 s_{n-1} + \cdots + p_n s_0]. \tag{2}$$

Hence Woronoi's definition of summability is identical to that used 17 years later by Nörlund,[2] and therefore includes as special cases those of Cesàro (C_α) and Riesz $R(n, \alpha)$. If we use the modern terminology we might restate the result above of Woronoi by saying that all his definitions of summability are *consistent*. The idea of the proof sketched by Woronoi (via Abel's theorem, as explained below) is precisely the same used for the proof of consistency by Zygmund[3] and Silverman-Tamarkin.[4] By Abel's theorem we mean the following statement:

---

[2] Sur une application des fonctions permutables, Lunds Univ. Årsskr., Avd. 2, 16 (1919), no. 3. Although various special cases of this definition have been used earlier, Nörlund's paper is, to the best of our knowledge, the first after that of Woronoi where the definition was stated in its complete generality. We use the term "Nörlund's transformations", although "Woronoi's transformations" would be more appropriate.

[3] O teorji srednich arythmetycznych, Mathesis Polska, 1 (1926), pp. 75–85, 119–129. Cf. also a correction, ibidem 5 (1930), p. 46.

[4] On the generalization of Abel's theorem for certain definitions of summability, Math. Zeit., 29 (1928), pp. 161–170, esp. pp. 169–170. The authors had no knowledge of the papers of Woronoi and Zygmund.

(A) *Whenever a sequence* $\{s_n\}$ *is such that the sequence* $\{\sigma_n\}$ *defined by* (2) *tends to a finite limit* σ *as* $n \to \infty$, *then the power series* $\sum_{n=0}^{\infty} u_n x^n$, *where* $u_0 + u_1 + \cdots + u_n = s_n$, *has a non-vanishing radius of convergence and defines an element of an analytic function* $u(x)$ *which tends to* σ *when* $x \to 1$ *along the radius* $0-1$.

Note that this statement does not exclude the possibility of singularities of $u(x)$ in the unit circle, or even on the segment $0-1$; only the portion of this segment sufficiently near to 1, $x<1$, must be free from singularities. To get Woronoi's notation we have only to replace x by e^{-x}. A brief analysis of the proof given by Silverman-Tamarkin shows that Woronoi's hypotheses can be made less restrictive. Indeed, we have formally

$$(3) \qquad \sum_{n=0}^{\infty} \sigma_n P_n x^n = \sum_{n=0}^{\infty} s_n x^n \sum_{n=0}^{\infty} p_n x^n = \sum_{n=0}^{\infty} u_n x^n \sum_{n=0}^{\infty} P_n x^n,$$

so that, if we set

$$(4) \qquad P(x) \equiv \sum_{n=0}^{\infty} P_n x^n, \quad \omega(x) \equiv \sum_{n=0}^{\infty} \sigma_n P_n x^n,$$

it follows

$$(5) \qquad u(x) \equiv \sum_{n=0}^{\infty} u_n x^n = \frac{\omega(x)}{P(x)}.$$

There is no loss of generality if we assume $p_0 = P_0 = 1$. Then (2), being written in the form

$$P_n \sigma_n = p_0 s_n + p_1 s_{n-1} + \cdots + p_n s_0$$

shows that we can take for $\{\sigma_n\}$ an arbitrary convergent sequence. In the particular case $\sigma_0 = 1$, $\sigma_n = 0$, $n = 1, 2, \cdots$, (5) reduces to $u(x) = 1/P(x)$, $P(x) = 1/u(x)$. Since, in this case, $u_0 = s_0 = \sigma_0 = 1$, it follows at once that, if (A) holds, the power series $P(x)$ must have a non-vanishing radius of convergence. We shall introduce, however, a more restrictive assumption, namely that the radius of convergence of $P(x)$ is equal to or greater than 1.

We shall designate by $\mathfrak{N}$ the corresponding class of Nörlund's transformations, using the notation $\mathfrak{N}^a$ for the subclass of $\mathfrak{N}$ for which Abel's theorem (A) holds. The notation $\mathfrak{N}^w$ will be used to designate the subclass of $\mathfrak{N}$ for which conditions (i), (ii) of Woronoi are satisfied. $\mathfrak{N}^r$ will designate the class of all regular Nörlund's transformations, $\mathfrak{N}^{rr}$ the subclass of $\mathfrak{N}^r$ of transformations with real coefficients. According to Woronoi, $\mathfrak{N}^w \subset \mathfrak{N}^a$, and it will be shown later that $\mathfrak{N}^{rr} \subset \mathfrak{N}^a$, while $\mathfrak{N}^{rr}$ and $\mathfrak{N}^w$ overlap. As it is shown in the following note by R. E. Gilman, $\mathfrak{N}^r$ and $\mathfrak{N}^a$ also overlap.

2. We shall now prove the following theorem.

(B) *In order that a Nörlund transformation of the class* $\mathfrak{N}$ *would belong to* $\mathfrak{N}^a$ *it is necessary and sufficient that* (i′) *there exist a number* ε, $0<\varepsilon<1$ *and a positive constant* M *such that*

$$\frac{1}{|P(x)|}\sum_{n=0}^{\infty}|P_n|x^n \leqq M, \qquad 1-\varepsilon \leqq x<1, \tag{6}$$

and (ii′) *the series* $P(1)$ *diverges, in the sense that*

$$\overline{\lim_{n\to\infty}}\left|\sum_{\nu=0}^{n}P_\nu\right| = \infty. \tag{7}$$

Condition (7) *can be replaced by*

$$\lim_{x\to 1}|P(x)| = \infty. \tag{8}$$

Proof. If (A) holds then the function $1/P(x) = u(x)$ for the particular choice $\sigma_0 = 1$, $\sigma_n = 0$, $n = 1, 2, \cdots$, must be free from singularities on $1-\varepsilon \leqq x<1$; moreover, since in this case $\sigma_n \to \sigma = 0$, we must have $u(x)\to 0$, or $|P(x)|\to\infty$, as $x\to 1$ along this segment. An easy application of Abel's partial summation formula shows that

$$\overline{\lim_{x\to 1}}|P(x)| \leqq \text{l. u. b.}_{n=0,1,2,\cdots}\left|\sum_{\nu=0}^{n}P_\nu\right|. \tag{9}$$

Hence, if (8) is satisfied, the right-hand member of (9) cannot be finite, which is equivalent to (7). On the other hand, if (6) and (7) are satisfied, the numerator in (6) tends to ∞ as $x\to 1$, which implies (8). Hence the set of conditions (6), (7) is equivalent to the set (6), (8).

Now it is obvious that the statement of Theorem (A) is equivalent to the statement that the transformation defined by

$$t(x) \equiv \sum_{k=0}^{\infty} b_k(x)\sigma_k, \quad b_k(x) = \frac{P_k x^k}{P(x)} \tag{10}$$

is regular in the sense of Silverman-Toeplitz theory. The conditions of regularity of such transformations are well known (cf. Silverman-Tamarkin, loc. cit.). In the present case they reduce to (6), (7), or to the equivalent set of conditions (6), (8).

It should be observed that all the transformations satisfying condition (ii) of Woronoi belong to $\mathfrak{N}$, while (i) implies (i′). If we agree to interpret Woronoi's "arithmetical divergence" so as to exclude the case of convergence, then it follows that $\mathfrak{N}^w \subset \mathfrak{N}^a$. The example $P_0 = 1$,

$P_n = n^{\log n}$, $n = 1, 2, \cdots$ shows that $\mathfrak{N}^w$ is a proper subclass of $\mathfrak{N}^a$ since the corresponding transformations $\subset \mathfrak{N}^a$ but do not $\subset \mathfrak{N}^w$.

Since Abel's theorem holds for all the transformations of $\mathfrak{N}^a$ it is plain that all these transformations are consistent because the limit σ, if it exists at all, must coincide with Abel's limit of the power series $u(x)$, which does not depend on the choice of the transformation. This is precisely what Woronoi has proved for the class $\mathfrak{N}^w$. It would be of interest to know whether there exist Nörlund's transformations which do not satisfy (i'), (ii'), but which are still consistent with all the transformations of $\mathfrak{N}^a$; and others which do not belong to $\mathfrak{N}$ but admit of Abel's theorem as stated in (A).

3. The conditions of regularity of a Nörlund's transformation are expressed by

$$1+\sum_{\nu=1}^{n} |P_\nu - P_{\nu-1}| = O(P_n), \; P_0 = 1 \tag{11}$$

and

$$\frac{p_n}{P_n} = \frac{P_n - P_{n-1}}{P_n} \to 0 \text{ or } \frac{P_{n-1}}{P_n} \to 1 \text{ as } n \to \infty. \tag{12}$$

In this case the radius of convergence of $P(x)$ is equal to 1, so that $\mathfrak{N}^r \subset \mathfrak{N}$.

(C) $\mathfrak{N}^{rr}$ *is a proper subclass of* $\mathfrak{N}^a$. *In other words, Abel's theorem holds for all regular Nörlund's transformations with real coefficients, while there exist transformations in* $\mathfrak{N}^a$ (*and indeed, in* $\mathfrak{N}^w$) *which are not regular.*

Proof. The proof of the second part of this theorem is furnished by the example

$$P_0 = P_1 = 1; \quad P_2 = P_3 = 2; \quad \cdots P_n = 2^{k-1} \text{ when } 2^{k-1} \leqq n < 2^k.$$

To prove the first part we have only to observe that (11) implies the existence of a constant $q > 0$ such that, for almost all n, $|P_n| > q$. Being combined with (12) this shows that almost all P_n are of the same sign. The verification of conditions (6), (7) now does not present any difficulty.

An analogous argument yields the truth of the theorem (C) *for the slightly more general class of regular Nörlund's transformations with complex coefficients, almost all of which are in a sector depending on the transformation, with the vertex at* 0 *and of an angle* $< \pi$.

On the basis of regularity conditions (11), (12) it can be proved that this is the case for all transformations of $\mathfrak{N}^r$ with $\overline{\lim_{n\to\infty}} |P_n| < \infty$. Indeed, on setting

$$P_n = R_n e^{i\varphi_n}, \qquad \varphi_n - \varphi_{n-1} = \delta_n,$$

we derive from (11), $\delta_n \to 0$. On the other hand there exists a sequence of indices $\{n_k\}$ such that

$$P_{n_k} = O(1), \qquad n_k \uparrow \infty.$$

This implies the convergence of the series $\sum_{n=1}^{\infty} |P_n - P_{n-1}|$. But, for almost all n, we have

$$|P_n - P_{n-1}| \geqq \left(4 R_n R_{n-1} \sin^2 \frac{\delta_n}{2}\right)^{1/2} \geqq \frac{q|\delta_n|}{2}.$$

Hence the series $\sum_{n=1}^{\infty} |\delta_n|$ also converges, thus yielding the existence of the sector in question. This sector may be selected even so as to have an arbitrarily small angle.[5] The existence of the sector in question is derived still simpler, according to a remark by Dr. R. Agnew, from the fact that the convergence of the series $\sum_{n=1}^{\infty} |P_n - P_{n-1}|$ implies the existence of $\lim_{n\to\infty} P_n$.

4. The only place in the literature (not counting a short review in the Jahrbuch für Fortschritte der Mathematik) where the note of Woronoi is mentioned at all, so far as we are aware, is the recent interesting monograph by E. Kogbetliantz.[6] The author states there: «En 1901, Woronoi a fait voir que pour les suites s_n à croissance finie le procédé (C, δ) est equivalent à tout procédé (Φ, δ) defini par

$$(\Phi, \delta) - \lim_{n\to\infty} \text{gén}\, s_n = \lim_{n\to\infty} \frac{1}{\Phi(n)} \sum_{0}^{n} \Phi(n-m)\, u_m = \lim_{n\to\infty} s_n^{(\delta)},$$

$\Phi(x)$ étant positive et monotone, $\Phi(x) \to \infty$ pour $x \to \infty$, pourvu que la limite $\lim n^{-(\delta+1)} \sum_{0}^{n} \Phi(m)$ existe et soit finie. Dans ces conditions la somme généralisée $-(\Phi, \delta)$[7] est indépendante de la fonction particulière $\Phi(x)$ employée.

Pour $\Phi(x) = x^{\delta}$ on a la suite $\{C_n^{(\delta)}\}$ où

$$C_n^{(\delta)} = \sum_{m<n} \left(1 - \frac{m}{n}\right)^{\delta} u_m$$

[5] As it is shown in the following note by R. E. Gilman, this result is "the best possible" in the sense that theorem (C) fails if sectors of an angle $\geqq \pi$ are allowed. There exist therefore regular transformations of $\mathfrak{R}^r$ which do not $\subset \mathfrak{R}^a$.

[6] Sommation des séries et intégrales divergentes par les moyennes arithmétiques et typiques, Mémorial des Sciences Mathématiques, Fasc. LI (1931), p. 52.

[7] We correct an obvious misprint $\Phi(x, s)$ in the text.

et $\lim_{n\to\infty} C_n^{(\delta)} = \lim_{n\to\infty} s_n^{(\delta)}$ si l'un des deux membres existe.[8] Le procédé (x^δ, δ) est un des procédés de Riesz, connus sous le nom des moyennes typiques et definis par rapport à une suite réelle, positive et monotone $\{\lambda_n\}$ telle que $\lambda_n \to \infty$ pour $n \to \infty$.» ··· «Les moyennes arithmétiques correspondent au type $\lambda_n = n$ et les procédés (C, δ) et (R, n, δ) sont équivalent, ce qui est le resultat de Woronoi, retrouvé par Riesz.»

If the reader compares this statement with the translation above of Woronoi's note, he will easily convince himself that neither Woronoi's hypotheses, nor results are stated here correctly. The only thing that can be inferred from Woronoi's note is that he had proved the *consistency* of the transformations in question whence the consistency of (C, δ) and (R, n, δ) immediately follows. The *equivalence* of these transformations is an entirely different matter. There is no proof of equivalence of any classes of transformations in Woronoi's note, and the result of Riesz does not follow from the result of Woronoi.

[8] The following example, due to Dr. R. Agnew, shows that this statement is incorrect. The sequence $(1, -1, 1, -1, \cdots)$ is summable to 0 by $(C, 1)$. It is not summable to any limit by $(\Phi, 1)$ where $\Phi(n) = n+1$, or $n+\frac{3}{2}$ according as n is odd or even.

BROWN UNIVERSITY.

A REMARK ON NÖRLUND'S METHOD OF SUMMATION.[1]

By R. E. Gilman.

In the preceding article[2] the notation of which will be used throughout the present note, it has been shown by Tamarkin that the class of regular Nörlund transformations whose coefficients are complex, but where, for each individual transformation, almost all are situated in a sector with vertex at O and of angle $<\pi$, is a sub-class $\mathfrak{N}^a$; in other words, Abel's theorem holds for all such transformations. The purpose of the present note is to show that this statement becomes untrue if sectors of angles $\geqq \pi$ are allowed, so that the result above appears as the best possible in a certain sense. To accomplish this it will be sufficient to exhibit an example of a power series

$$P(x) \equiv \sum_{n=0}^{\infty} P_n x^n \tag{1}$$

with the following properties:

i. $$\frac{P_{n+1}}{P_n} \to 1 \text{ as } n \to \infty .$$

ii. $$\sum_{\nu=1}^{n} |P_\nu - P_{\nu-1}| = O(P_n).$$

iii. The ratio

$$\frac{P^*(x)}{|P(x)|} \tag{2}$$

where

$$P^*(x) \equiv \sum_{n=0}^{\infty} |P_n| x^n \tag{3}$$

does not remain bounded as $x \to 1$ along the segment $(0,1)$.[3]

The series in question will be constructed as follows. We start with the expansion

$$(1-x)^{-2} = 1+2x+3x^2+\cdots+(n+1)x^n+\cdots. \tag{4}$$

Next we change the signs of the coefficients to make them alternate in successive groups of 1, 2, 4, $\cdots$, 2^k, $\cdots$ terms, thus obtaining

$$g(x) \equiv 1+\sum_{k=1}^{\infty} (-1)^k \sum_{\nu=1}^{2^k} (2^k+\nu-1)\, x^{2^k+\nu-2}. \tag{5}$$

[1] Received March 5, 1932.

[2] G. Woronoi, Extension of the notion of the limit of the sum of terms of an infinite series, translated by J. Tamarkin, these Annals, pp. 422–428.

[3] This will be selfunderstood when we say $x \to 1$, and in general, only positive real values of $x<1$ will be considered in our discussion.

Finally we introduce the notation

$$\theta(k,\nu) = \begin{cases} \exp i\pi\left(1+\frac{\nu}{k}\right), & k \text{ odd}, \\ \exp i\pi\left(1-\frac{\nu}{k}\right), & k \text{ even}, \end{cases} \tag{6}$$

and in each group of 2^k terms modify the coefficients of the k first terms of the group on multiplying them by $\theta(k,\nu)$. The resulting series

$$\begin{aligned} P(x) &\equiv \sum_{n=0}^{\infty} P_n x^n \\ &= 1+\sum_{k=1}^{\infty}(-1)^k\Big[\sum_{\nu=1}^{k}\theta(k,\nu)\,(2^k+\nu-1)\,z^{2^k+\nu-2} \\ &\qquad + \sum_{\nu=k+1}^{2^k}(2^k+\nu-1)\,z^{2^k+\nu-2}\Big] \end{aligned} \tag{7}$$

will be shown to satisfy all our requirements i.–iii.[4]

We first observe that all P_n are obtained from the coefficients of (4) by multiplying them by factors of the type $e^{i\alpha}$. Hence

$$|P_n| = n+1, \quad P^*(x) = (1-x)^{-2}. \tag{8}$$

Consequently

$$\left|\frac{P_{n+1}}{P_n}\right| \to 1; \quad \left|\arg\left(\frac{P_{n+1}}{P_n}\right)\right| \leqq \frac{\pi}{k} \to 0 \text{ as } n,\ k \to \infty.$$

This establishes property i.

To prove ii. let

$$n = 2^k+m < 2^{k+1}, \quad 0 \leqq m < 2^k. \tag{9}$$

We write

$$\sum_{\nu=1}^{n}|P_\nu - P_{\nu-1}| \equiv \sum\nolimits_1 + \sum\nolimits_2, \tag{10}$$

where $\sum_1$ is that part of the summation extended over the values of ν for which

$$2^\mu - 1 \leqq \nu \leqq 2^\mu + \mu - 1, \qquad \mu = 1,\ 2,\ \cdots,\ k. \tag{11}$$

In the remaining part of the summation, $\sum_2$, we have

$$|P_\nu - P_{\nu-1}| = 1,$$

whence

$$\sum\nolimits_2 \leqq n = O(P_n). \tag{12}$$

As for $\sum_1$, if ν satisfies (11) with a fixed μ, it is readily found

$$|P_\nu - P_{\nu-1}| = \nu\left|1-\frac{\nu+1}{\nu}\exp\left((-1)^\mu\frac{\pi i}{\mu}\right)\right| \leqq (\nu+1)\frac{\pi}{\mu}.$$

[4] Note also that all the coefficients P_n lie in a sector of angle π, namely the upper half plane.

Accordingly the summation extended over all ν's of this group (μ in number except, perhaps the last group, $\mu = k$, when the number of ν's might be less than k) will be less than the product of μ times the maximum term of the group, which itself does not exceed $\frac{\pi 2^{\mu+1}}{\mu}$. Hence

$$\sum_1 < \sum_{\mu=1}^{k} \pi 2^{\mu+1} < \pi 2^{k+2} = O(P_n), \tag{13}$$

and the property ii. follows if we combine (12) and (13).

The proof of iii. is more complicated. In view of (8) this property will be established if we can show the existence of a sequence $\{x_j\}$ such that the product

$$P(x)(1-x)^2 \to 0 \text{ as } x_j \to 1. \tag{14}$$

Set

$$P(x) = g(x) + h(x) \tag{15}$$

where

$$h(x) = P(x) - g(x) = \sum_{k=1}^{\infty} (-1)^k \sum_{\nu=1}^{k} \{\theta(k,\nu) - 1\}(2^k + \nu - 1) x^{2^k+\nu-2}. \tag{16}$$

It will be sufficient for our purposes to have a crude estimate of the order of magnitude of $h(x)$ as $x \to 1$. Since $|\theta(k,\nu) - 1| \leqq 2$, we have at once

$$|h(x)| \leqq 2 \sum_{k=1}^{\infty} \sum_{\nu=1}^{k} (2^k + \nu - 1) x^{2^k+\nu-2}. \tag{17}$$

A simple computation shows that, for a given x, $0 < x < 1$, the expression $(n+1)x^n$ reaches its maximum $= O[(1-x)^{-1}]$ when $n = O[x(1-x)^{-1}]$. Let now

$$\sum_{k=1}^{\infty} \sum_{\nu=1}^{k} = \sum_{k=1}^{M} \sum_{\nu=1}^{k} + \sum_{k=M+1}^{\infty} \sum_{\nu=1}^{k} \equiv S_N + R_N, \tag{18}$$

where $N = 2^M$ is the highest power of 2 less than $\frac{\log(1-x)}{(x-1)}$. It is clear that R_N is less than the remainder after N terms of the series (4), which is $O(1)$, for N chosen as specified. As for S_N, if we replace each term by the value of the largest term namely $O[(1-x)^{-1}]$, we get

$$S_N \leqq \sum_{k=1}^{M} \sum_{\nu=1}^{k} O[(1-x)^{-1}] = O(M^2)\, O[(1-x)^{-1}] = O\left[\frac{\log^2(1-x)}{1-x}\right],$$

whence

$$h(x) = O\left[\frac{\log^2(1-x)}{1-x}\right] \text{ as } x \to 1. \tag{19}$$

We proceed quite differently with $g(x)$. We have

$$
\begin{aligned}
g(x) &= 1+\sum_{k=1}^{\infty}\sum_{\nu=1}^{2^k}(-1)^k(2^k+\nu-1)x^{2^k+\nu-2} \\
&= \frac{d}{dx}\left[-1+x+\sum_{k=1}^{\infty}\sum_{\nu=1}^{2^k}(-1)^k x^{2^k+\nu-1}\right] \qquad (20) \\
&= \frac{d}{dx}\left\{\left[-1+2\sum_{k=0}^{\infty}(-1)^k x^{2^k}\right](1-x)^{-1}\right\} \\
&= -2\left[(1-x)f'(x)+f(x)\right](1-x)^{-2} = -2\varphi(x)(1-x)^{-2},
\end{aligned}
$$

where

$$
f(x) \equiv \tfrac{1}{2}-\sum_{k=0}^{\infty}(-1)^k x^{2^k} \qquad (21)
$$

and

$$
\varphi(x) \equiv (1-x)f'(x)+f(x). \qquad (22)
$$

It follows from (21) that

$$
f(x)+f(x^2) = 1-x \qquad (23)
$$

and by differentiation

$$
f'(x)+2xf'(x^2) = -1. \qquad (24)
$$

From (22), (23), (24) we derive

$$
\varphi(x)+\varphi(x^2) = (1-x)^2 f'(x^2). \qquad (25)
$$

Take an arbitrary sequence E of values of $x\to 1$. If, on E, $f'(x)$ is not bounded then from (24) $f'(x)$ and $f'(x^2)$ assume opposite signs so that $f'(x)$ vanishes between x and x^2. Thus a set E' of values will always exist on which $f'(x)$ is bounded. For x^2 on E' the right member of (25)$\to 0$, so that $\varphi(x)+\varphi(x^2)\to 0$. Then either $\varphi(x)\to 0$, or else $\varphi(x)$ and $\varphi(x^2)$ assume opposite signs and, by continuity, $\varphi(x)$ vanishes for some x' between x and x^2. In either case there exists a sequence $\{x_j\}$ of values of $x\to 1$ on which $\varphi(x)\to 0$.

If we restrict x to assume only the values belonging to this sequence, it follows from (19) and (20) that

$$
P(x)(1-x)^2 = (1-x)^2[g(x)+h(x)] = -2\varphi(x)+O[(1-x)\log^2(1-x)]\to 0
$$

which is precisely what we wanted to show.

BROWN UNIVERSITY.

A CLASS OF FUNCTIONS CONTINUOUS BUT NOT ABSOLUTELY CONTINUOUS.*

By R. E. Gilman.

1. **Introduction.** In a recent paper,[1] Hille and Tamarkin have set forth a collection of interesting properties of the well known function of Cantor which is continuous and monotone but not absolutely continuous. The following pages are devoted to a discussion of a more general class of functions, containing the one just mentioned as a special case. Every property (except for trifling modifications) noted by Hille and Tamarkin for the special case, is true for all the functions of this more general class, but not to be tedious, this statement will be justified only in part. Our attention will be centered mainly on a study of the behavior of the derivatives, a topic cited by the above authors as having been considered but not solved.[2]

2. **The perfect set $P_{\alpha\beta}$.** Represent the numbers of the closed interval [0,1] in an arithmetic scale of radix α. Thus

$$x = \frac{a_1}{\alpha} + \frac{a_2}{\alpha^2} + \cdots + \frac{a_n}{\alpha^n} + \cdots \tag{1}$$

where α is a positive integer greater than 2, and $a_1, a_2, \cdots, a_n, \cdots$ are integers such that

$$0 \leqq a_i < \alpha \qquad (i = 1, 2, \cdots). \tag{2}$$

Hereafter a_ν will be referred to as the ν-th digit in the representation of x.

Let $\alpha - 1$ be represented as the product of two integers as follows

$$\alpha - 1 = q(\beta - 1), \quad q \geqq 2, \quad \beta \geqq 2, \tag{3}$$

which is evidently always possible. It will be assumed that a definite choice of α, β, and q is made so that they are hereafter to be regarded as fixed.

We define the set $P_{\alpha\beta} \equiv P$ as consisting of all the points x of [0,1], for which all the digits in a representation (1) are integral multiples of q.

* Received August 18, 1931.

[1] Remarks on a known example of a monotone continuous function. American Math. Monthly, 36 (1929), pp. 255-264.

[2] Loc. cit., footnote on p. 258.

The complementary set $CP \equiv C$ will comprise all the points of [0,1] for which no representation in the scale α has every digit a multiple of q.

The set P will be subdivided further into three subsets, P^0, P^+, P^-, where

P^0 comprises all the points of P having an infinite number of digits different from zero and from $\alpha-1$;

P^+ comprises all the points of P with only a finite number of digits different from zero;

P^- comprises all the points of P with only a finite number of digits different from $\alpha-1$.

It should be observed that P_{32} reduces to the classical Cantor set considered by Hille and Tamarkin.

The following properties of the sets C and P are easily derived.

i. The set C is open and consists of all the open (non-overlapping) intervals (ξ', ξ'') with the end points

$$(4)\qquad \begin{aligned} \xi' &= \frac{a_1'}{\alpha} + \cdots + \frac{a_{k-1}'}{\alpha^{k-1}} + \frac{a_k' - q + 1}{\alpha^k}, \\ \xi'' &= \frac{a_1'}{\alpha} + \cdots + \frac{a_{k-1}'}{\alpha^{k-1}} + \frac{a_k'}{\alpha^k}, \end{aligned}$$

where $a_1', a_2', \cdots, a_{k-1}', a_k'$ are integral multiples of q and $a_k' \neq 0$.

An interval of this type will be designated as an interval of order k.

ii. The sum of all the lengths of all the intervals of all orders $k = 1, 2, \cdots$ is equal to one, whence meas. $C = 1$.

iii. The set P is perfect, non-dense in [0, 1], and of measure zero. The intervals of D are precisely the contiguous ("black") intervals of P the points of P^+, P^- being, respectively, the right and left end points of (ξ', ξ'').

It should be noted first that some numbers x admit of ambiguous representation of type (1), for,

$$\begin{aligned} \frac{a+1}{\alpha^n} &= \frac{a}{\alpha^n} + \frac{\alpha-1}{\alpha^{n+1}} + \cdots, \quad 0 \leqq a < \alpha - 1, \\ \frac{a}{\alpha^n} &= \frac{a-1}{\alpha^n} + \frac{\alpha-1}{\alpha^{n+1}} + \cdots, \quad 0 < a \leqq \alpha - 1. \end{aligned}$$

Hence a point x will be classified to belong to P, P^+, P^-, if at least one of its representations (1) satisfies the corresponding requirements.

Let now $x \subset C$. We shall have then a representation (1) with

$$a_i \equiv 0 \ (q), \quad i = 1, 2, \cdots, k-1; \quad a_k \not\equiv 0 \ (q),$$

not all $a_{k+1}, \cdots$ being equal to zero (or to $\alpha - 1$) when $a_k \equiv 1$ (or $\equiv -1$). It is obvious under these conditions that x is an interior point of an

interval (ξ', ξ'') of order k, namely that for which a'_k is the smallest multiple of q greater than a_k, while $a'_i = a_i$ $(i = 1, 2, \cdots, k-1)$. It is also plain that all the interior points of $(\xi', \xi'') \subset C$, which establishes the property i. To prove property ii. we observe that the length of any interval of order k is $\frac{q-1}{\alpha^k}$. If now $k = 1$, a'_1 may have any of the values $q, 2q, \cdots, (\beta-1)q$, hence the sum of lengths of all the intervals of order 1 is $(\beta-1)\frac{q-1}{\alpha}$. If $k>1$, the digits $a'_1, a'_2, \cdots, a'_{k-1}$ may assume all the values $0, q, 2q, \cdots, (\beta-1)q$, while a'_k may assume the same values with exception of 0. Hence the sum of lengths of all the intervals of order k is

$$\beta^{k-1}(\beta-1)(q-1)/\alpha^k,$$

and the sum of lengths of all the intervals of C, that is meas. C, is

$$\sum_{k=1}^{\infty} \beta^{k-1}(\beta-1)(q-1)\alpha^{-k} = \frac{(\beta-1)(q-1)}{\alpha} \frac{1}{1-\frac{\beta}{\alpha}} = 1.$$

This proves property ii. Property iii. is now obvious. The following property will be of great use in our subsequent discussion; the proof is omitted.

iv. The transformation $x' = 1-x$ leaves the sets C, P, P^0 invariant, while it interchanges the sets P^+, P^-.

3. The function $\omega_{\alpha\beta}(x)$. We shall define now a function $\omega_{\alpha\beta}(x) \equiv \omega(x)$ which reduces to that considered by Hille and Tamarkin in the special case $\alpha = 3$, $\beta = 2$.

Let x have a representation of the type (1). We set

$$\omega(x) = \frac{b_1}{\beta} + \frac{b_2}{\beta^2} + \cdots + \frac{b_n}{\beta^n} + \cdots \tag{5'}$$

when all the digits $a_i \equiv 0\ (q)$ and

$$a_i = q b_i \qquad (i = 1, 2, \cdots), \tag{6'}$$

while

$$\omega(x) = \frac{b_1}{\beta} + \frac{b_2}{\beta^2} + \cdots + \frac{b_{n-1}}{\beta^{n-1}} + \frac{b_n}{\beta^n}, \quad b_n = \left[\frac{a_n}{q}\right] + 1 \tag{5''}$$

when

$$a_i = q b_i\ (i = 1, 2, \cdots, n-1) \text{ but } a_n \not\equiv 0\ (q). \tag{6''}$$

A moment's consideration will show that $\omega(x)$ thus defined is a single-valued function of x, although in some cases x admits of two representations of type (1). Various simple properties of $\omega(x)$ are stated in the following

THEOREM 1. i. *The function* $\omega(x)$ *does not decrease and varies from* $\omega(0) = 0$ *to* $\omega(1) = 1$ *as* x *ranges over* $[0, 1]$.

ii. $\omega(x)$ *assumes equal values at the end points of each interval* (ξ', ξ'') *contiguous to the set* P, *these intervals thus being intervals of constancy of* $\omega(x)$.

iii. $\omega(x)$ *satisfies the functional equation*

$$\omega(x) + \omega(1-x) = 1. \tag{7}$$

iv. $\omega(x)$ *is continuous but not absolutely continuous.*

Proof. To prove i. we have to show that if x and x' are any two points of $[0, 1]$ and $x > x'$, then $\omega(x) \geqq \omega(x')$. Let

$$x = \frac{a_1}{\alpha} + \cdots + \frac{a_{k-1}}{\alpha^{k-1}} + \frac{a_k}{\alpha^k} + \cdots,$$

$$x' = \frac{a_1}{\alpha} + \cdots + \frac{a_{k-1}}{\alpha^{k-1}} + \frac{a'_k}{\alpha^k} + \cdots$$

be the representations of x, x' so that a'_k is the first digit of x' distinct from the corresponding digit of x. Then

$$a_k > a'_k,$$

the possibility

$$a_k - a'_k = 1, \qquad a_{k+1} = a_{k+2} = \cdots = 0, \qquad a'_{k+1} = a'_{k+2} = \cdots = \alpha - 1$$

being excluded.

If there is at least one among the digits $a_1, \cdots, a_{k-1}$ that is $\not\equiv 0\ (q)$, the equality $\omega(x) = \omega(x')$ follows from the definition (5″). On the other hand, if all the digits $a_1, \cdots, a_k, a'_k$ are $\equiv 0\ (q)$, then

$$b_i = b'_i \quad (i = 1, 2, \cdots, k-1), \qquad b_k > b'_k.$$

If $a_k \not\equiv 0\ (q)$, $a'_k \not\equiv 0\ (q)$, we have the same situation, except that $>$ in the last relation might be replaced by $\geqq$. Since in this case both sequences (b_n), (b'_n) terminate at $n = k$, we have, in both cases, $\omega(x) \geqq \omega(x')$. If $a_k \equiv 0\ (q)$ while $a'_k \not\equiv 0\ (q)$, the sequence (b'_n) terminates at $n = k$ and $b'_k \leqq b_k$. Finally, if $a_k \not\equiv 0\ (q)$ while $a'_k \equiv 0\ (q)$, we have $b_k > b'_k$, so that, in both cases, $\omega(x) \geqq \omega(x')$. The relations $\omega(0) = 0$, $\omega(1) = 1$ follow immediately from the definition of $\omega(x)$, as well as the property ii. of the theorem.

To prove iii. we observe that if $x \subset P$, then $x' = 1 - x \subset P$. Let (a_n), (a'_n) be the corresponding sequences of digits in the representations (1) of x, x'. We have

$$a_i = q\, b_i, \qquad a'_i = (\alpha - 1) - a_i = q(\beta - 1 - b_i), \qquad b'_i = (\beta - 1) - b_i,$$

$$\omega(x) + \omega(x') = \sum_{i=1}^{\infty} b_i\, \beta^{-i} + \sum_{i=1}^{\infty} b'_i\, \beta^{-i} = \sum_{i=1}^{\infty} (\beta - 1)\, \beta^{-i} = 1.$$

If, however, $x \subset C$, there will be a positive integer k such that the relations above hold for $i \leqq k-1$, while

$$b_k = \left[\frac{a_k}{q}\right]+1, \quad b_k' = \left[\frac{a_k'}{q}\right]+1, \quad a_k + a_k' = \alpha - 1, \quad b_k + b_k' = \beta - 1.$$

The sequences (b_n), (b_n') now terminate at $k = n$, and we get again $\omega(x)+\omega(x') = 1$. The continuity of $\omega(x)$ will follow from the more precise Theorem 2 below. To prove that $\omega(x)$ is not absolutely continuous we have only to observe that, on one hand,

$$\omega(1) - \omega(0) = 1,$$

while, on the other hand, the derivative $\omega'(x) = 0$ at all the points interior to the intervals of constancy of $\omega(x)$, that is at all points $x \subset C$. Since meas. $C = 1$ this implies

$$\int_0^1 \omega'(x)\,dx = 0 \neq \omega(1) - \omega(0).$$

Hence $\omega(x)$ is not absolutely continuous.

THEOREM 2. *The function $\omega(x)$ satisfies a Lipschitz condition of order*

$$\mu = \frac{\log \beta}{\log \alpha}, \quad \beta = \alpha^{\mu}, \tag{8}$$

the Lipschitz coefficient being not greater than

$$c = \beta(q-1)^{-\mu}. \tag{9}$$

This result is the "best possible" in the sense that the inequality

$$|\omega(x) - \omega(x')| \leqq c_0 |x - x'|^{\mu_0}$$

does not hold for unrestricted pairs of values (x, x') on $[0,1]$, even for those with $x'-x \to 0$, if either $\mu_0 > \mu$ or $\mu_0 = \mu$ but $c_0 < c$.

Proof. There is no restriction of generality in assuming $x > x'$. First suppose $x \subset P$, $x' \subset P$. We have then

$$x = \sum_{i=1}^{\infty} a_i \alpha^{-i}, \quad x' = \sum_{i=1}^{\infty} a_i' \alpha^{-i}, \quad a_i = q b_i, \quad a_i' = q b_i',$$

$$\omega(x) = \sum_{i=1}^{\infty} b_i \beta^{-i}, \quad \omega(x') = \sum_{i=1}^{\infty} b_i' \beta^{-i}.$$

Let a_k be the first digit of x which is different from the corresponding digit a_k' of x'. Then

$$a_k - a_{k'} \geqq q,$$

$$x - x' = (a_k - a_k')\alpha^{-k} + \sum_{i=k+1}^{\infty} a_i \alpha^{-i} - \sum_{i=k+1}^{\infty} a_i' \alpha^{-i} \geqq (q-1)\alpha^{-k},$$

and

$$\omega(x) - \omega(x') = (b_k - b_k')\beta^{-k} + \sum_{i=k+1}^{\infty} b_i \beta^{-i} - \sum_{i=k+1}^{\infty} b_i' \beta^{-i}$$
$$\leqq (\beta - 1)\beta^{-k} + \beta^{-k} = \beta^{1-k} = \beta\alpha^{-k\mu}.$$

Hence

$$0 \leqq \omega(x) - \omega(x') \leqq \beta\left(\frac{x-x'}{q-1}\right)^{\mu} = c(x-x')^{\mu}. \tag{10}$$

Let now either x, or x', or both, $\subset C$. Let ξ_2 be the left end point of the interval of constancy containing x, and ξ_1 the right end point of the interval of constancy containing x'. Then, by the preceding result,

$$\omega(x) - \omega(x') = \omega(\xi_2) - \omega(\xi_1) \leqq c(\xi_2 - \xi_1)^{\mu} \leqq c(x-x')^{\mu},$$

since ξ_1 and $\xi_2 \subset P$. The proof of Theorem 2 is now complete.[3]

4. **The incremental ratios and derivatives of $\omega(x)$.** We are now ready to study the derivatives of $\omega(x)$.

THEOREM 3. *At all the points of C, $\omega(x)$ has the unique derivative* 0. *At all the points of P^+,*

$$D^+\omega(x) = +\infty, \quad D^-\omega(x) = 0, \tag{11}$$

while at those of P^-,

$$D^+\omega(x) = 0, \quad D^-\omega(x) = +\infty, \tag{12}$$

with the usual notation $D^{\pm}$ for the derivatives on the right or on the left.

Proof. The statement of Theorem 3 concerning the set C has been proved already [Theorem 1, ii.]. Since the statement concerning P^- is derived from that concerning P^+ by the transformation $x' = 1-x$, only formulas (11) need to be proved. The second of formulas (11) is obvious since every point of P^+ is the right end point of an interval of constancy of $\omega(x)$. Let now

$$x = \sum_{i=1}^{n} a_i \alpha^{-i}, \quad a_i = q b_i$$

be a point of P^+ and $x' > x$, while $x' \to x$. There exists an integer $p > n$ such that

$$x < x_1 = x + \alpha^{-p-1} \leqq x' \leqq x_2 = x + \alpha^{-p}, \; p \to \infty \text{ as } x' \to x.$$

[3] Cf. the corresponding discussion of Hille and Tamarkin, loc. cit., pp. 259-260.

Then, by Theorem 1, i. and (5″),

$$\omega(x) < \omega(x_1) = \omega(x) + \beta^{-p-1} \leqq \omega(x') \leqq \omega(x_2) = \omega(x) + \beta^{-p},$$

whence

$$\alpha^{p+1}\beta^{-p} \geqq \frac{\omega(x') - \omega(x)}{x' - x} \geqq \alpha^p \beta^{-p-1} > \left(\frac{\alpha}{\beta}\right)^p \to \infty \text{ as } x' \to x,$$

which shows that $D^+\omega(x) = +\infty$.

The behavior of $\omega(x)$ at the points of the set P^0 is more interesting and more complicated. The following notation will prove useful.

Let x be a fixed point of P^0 and

$$x = \sum_{k=1}^{\infty} a_{\nu_k} \alpha^{-\nu_k}, \quad a_{\nu_k} = q b_{\nu_k} \tag{13}$$

the corresponding expansion. Without loss of generality we may here assume that

$$\nu_1 < \nu_2 < \cdots \nu_k < \cdots \to \infty, \tag{14}$$

$$\alpha_k \equiv a_{\nu_k} \neq 0, \qquad \beta_k \equiv b_{\nu_k} \neq 0, \tag{15}$$

while the possibility

$$a_{\nu_k} = \alpha - 1, \quad \nu_{k+1} - \nu_k = 1, \quad k \geqq k_0 \geqq 1$$

is excluded.

By x' we designate the variable point which ultimately will $\to x$, and by

$$Q(x', x) \equiv Q(x') \equiv \frac{\omega(x') - \omega(x)}{x' - x} \tag{16}$$

the corresponding difference quotient.

We set

$$x_n = \sum_{k=1}^{n} \alpha_k \alpha^{-\nu_k}, \quad y_n = \sum_{k=1}^{n-1} \alpha_k \alpha^{-\nu_k} + \frac{\alpha_n - q + 1}{\alpha^{\nu_n}}, \tag{17}$$

$$\delta_n = [x_{n-1}, x_n]. \tag{18}$$

In order to have some information concerning the distribution of the 0's and $(\alpha-1)$'s in the expansion of x, we observe that ν_n is the place of the n-th non zero digit of x, a_{ν_n}, and denote by

$$s_n = \nu_{n+1} - \nu_n - 1 \tag{19}$$

the number of zero digits that are between a_{n_ν} and the next non zero digit $a_{n_{\nu+1}}$. We shall be interested in the ratios

$$r_n = \frac{s_n}{\nu_n} \qquad (n = 1, 2, \cdots). \tag{20}$$

We shall use (ε_n), (η_n) as a generic notation for sequences of positive numbers such that

$$\varepsilon_n \to 0,\ \eta_n \to 0,\ \text{while } \nu_n \varepsilon_n \to \infty,\ \nu_n \eta_n < M, \tag{21}$$

where M is a positive constant. Finally, by ν_n^*, s_n^*, r_n^*, ε_n^*, η_n^* we designate the quantities which should replace the corresponding quantities ν_n, s_n, r_n, ε_n, η_n when "zero digit" in the above definitions is replaced by "$\alpha - 1$ digit". To abbreviate, let λ denote the positive constant

$$\lambda = \frac{1}{\mu} - 1 = \frac{\log \alpha}{\log \beta} - 1. \tag{22}$$

With this notation we have the following lemmas.

LEMMA 1. *When $x' \to x$ on the sequence (x_n), the difference quotient*

$$Q(x_n) > \frac{\beta_{n+1}}{\alpha_{n+1} + 1} \left(\frac{\alpha}{\beta}\right)^{\nu_{n+1}} \to \infty. \tag{23}$$

The proof follows immediately from the relations

$$\omega(x) - \omega(x_n) = \sum_{k=n+1}^{\infty} \beta_k \beta^{-\nu_k} > \beta_{n+1} \beta^{-\nu_{n+1}},$$

$$x - x_n = \sum_{k=n+1}^{\infty} \alpha_k \alpha^{-\nu_k} < (\alpha_{n+1} + 1) \alpha^{-\nu_{n+1}},$$

$$\nu_{n+1} \to \infty, \qquad \left(\frac{\alpha}{\beta}\right)^{\nu_{n+1}} \to \infty.$$

LEMMA 2. *At $x' = y_n$ the difference quotient $Q(x')$ satisfies*

$$Q(y_n) < \frac{\beta_{n+1} + 1}{q - 1} \alpha^{\nu_n} \beta^{-\nu_{n+1}}. \tag{24}$$

For, now we have

$$\omega(x) - \omega(y_n) = \sum_{k=n+1}^{\infty} \beta_k \beta^{-\nu_k} < (\beta_{n+1} + 1) \beta^{-\nu_{n+1}},$$

$$x - y_n = (q-1) \alpha^{-\nu_n} + \sum_{k=n+1}^{\infty} \alpha_k \alpha^{-\nu_k} > (q-1) \alpha^{-\nu_n},$$

whence (24) follows.

LEMMA 3. *Whenever $x' \subset \delta_n \equiv [x_{n-1}, x_n]$, we have*

$$Q(x') > \frac{\beta_{n+1}}{\alpha_n + 1} \alpha^{\nu_n} \beta^{-\nu_{n+1}} > \frac{1}{\alpha \beta} Q(y_n). \tag{25}$$

Proof. Since $\omega(x)$ is non-decreasing, whenever $x_{n-1} \leqq x' \leqq x_n$, we have

$$\omega(x_{n-1}) \leqq \omega(x') \leqq \omega(x_n)$$

and

$$\omega(x) - \omega(x') \geqq \omega(x) - \omega(x_n) > \beta_{n+1} \beta^{-\nu_{n+1}},$$

while

$$x - x' \leqq x - x_{n-1} < (\alpha_n + 1)\, \alpha^{-\nu_n}.$$

This yields the first part of (25), whence, by (24),

$$\frac{Q(x')}{Q(y_n)} > \frac{\beta_{n+1}(q-1)}{(\alpha_n+1)(\beta_{n+1}+1)} > \frac{1}{\alpha\beta}.$$

We come now to our main result:

THEOREM 4. *A necessary and sufficient condition that, at a point $x \subset P^0$, $\omega(x)$ have a derivative on the left is that*

$$Q(y_n) \to \infty \quad \text{as} \quad n \to \infty. \tag{26}$$

This derivative is equal to $+\infty$ if it exists at all.

Proof. The necessity of (26) follows from Lemma 1, since $y_n \to x$, while, for $x_n \to x$, $Q(x_n) \to \infty$. The sufficiency of (26) follows from Lemma 3 combined with the fact, that, for x' sufficiently near to x, there will exist an integer n such that $x' \subset \delta_n$, $n \to \infty$.

It is not difficult, by using the transformation $t = 1 - x$, to obtain the analogue of Theorem 4 for the derivative on the right at an $x \subset P^0$. Probably a better grasp of the true nature of the condition (26) can be obtained from the following useful specialization of Theorem 4.

THEOREM 5. *If $x \subset P^0$ and*

$$r_n \leqq \lambda - \varepsilon_n, \qquad n > N \tag{27}$$

for all values of n greater than some fixed constant N, the function $\omega(x)$ has a derivative on the left, equal to $+\infty$. If, however, we have

$$r_n \geqq \lambda - \eta_n \tag{28}$$

for infinitely many values of n, $\omega(x)$ has no derivative on the left.

The same is true of the derivative on the right, provided the conditions (27), (28) *are replaced respectively by*

$$r_n^* \leqq \lambda - \varepsilon_n^*, \qquad n > N^* \tag{27*}$$

and

$$r_n^* \geqq \lambda - \eta_n^*. \tag{28*}$$

Proof. It suffices to consider only the derivatives on the left. Those on the right can be treated by our customary transformation $t = 1 - x$. Since $y_n \subset \delta_n$, we have from (24) and (25),

$$\frac{\beta_{n+1}}{\alpha_n+1}\alpha^{\nu_n}\beta^{-\nu_{n+1}} < Q(y_n) < \frac{\beta_{n+1}+1}{q-1}\alpha^{\nu_n}\beta^{-\nu_{n+1}}.$$

Hence

$$Q(y_n) = c_n \alpha^r \beta^{-\nu_{n+1}} = c_n \beta^{\lambda\nu_n - s_n}, \tag{29}$$

where

$$\frac{1}{\alpha} \leqq c_n \leqq \beta, \tag{30}$$

so that c_n is bounded from infinity above and from zero below. If we write (29) in the form

$$Q(y_n) = c_n \beta^{\nu_n(\lambda - r_n)} \tag{31}$$

we see at once that $Q(y_n) \to \infty$ when (27) is satisfied, in view of (21), while $Q(y_n)$ can not $\to \infty$ when (28) is satisfied.

It is worth while to recapitulate the above results by classifying the various cases of the behavior of the derivatives of $\omega(x)$.

I. Unique derivative $(+\infty)$.
II. Derivative $(+\infty)$ on the right, no derivative on the left.
III. Derivative $(+\infty)$ on the left, no derivative on the right.
IV. No derivative on the right, nor on the left.
} $x \subset P^0$.

V. Derivative $(+\infty)$ on the right and (0) on the left. $x \subset P^+$.
VI. Derivative (0) on the right and $(+\infty)$ on the left. $x \subset P^-$.
VII. Unique derivative (0). $x \subset C$.

It should be observed that, as far as the types of the behavior of derivatives are concerned, the sets P^0, P^+, P^-, C are mutually exclusive, while the types I–IV actually occur, each on a non-denumerable sub-set of P^0.

We might say in conclusion that all the rational points $\subset P^0$ are of type I since in expansions (1) of such points the digits repeat periodically, so that both s_n and s_n^* are bounded and the ratios r_n, $r_n^* \to 0$. On the contrary, the numbers of Liouville, that are obtained from (13) for $\nu_n = n!$, are all of type II, since, for such numbers, $r_n \to \infty$ and $r_n^* \to 0$.

BROWN UNIVERSITY.

RELATIVE SUMMABILITY.[1]

By R. L. Jeffery.

1. **Introduction.** In a previous note[2] we discussed a problem in integration which involved non measurable sets. As one would expect under such circumstances, Lebesgue's theory proved inadequate for its solution. The solution was obtained by a device which we have since learned is none other than Hildebrandt's *relative summability.*[3] Relative summability is obviously well adapted to the handling of such problems, and for this reason seems to deserve more consideration than Hildebrandt gave to it. We show that it can be made the basis of a complete and satisfactory theory of integration which includes those of Lebesgue, of Pierpont, and of Young, and this without in any way featuring measurable sets. If S is any set we use outer measure, designated by the symbol $l\,S$, which is defined for every bounded set, and which is equivalent to measure when the sets involved happen to be measurable. We find, however, that there is no necessity for distinguishing between the two cases.

2. **Definitions and notation.** Let f be any bounded function defined on the bounded set S. Let $a_1, a_2, \cdots, a_n$ be a sub-division of the range of f, and s_i the part of S for which $a_{i-1} \leqq f < a_i$. If there exists an unique number I such that

$$\sum_{i=1}^{n} a_i\, l\, s_i$$

is arbitrarily near to I provided the norm of the sub-division is sufficiently small, we shall say that this number I is the Hildebrandt integral of f over S, denoted by the symbol $H\,t\int_S f\,d\,x$.

If S is an n-dimensional domain we are able to show that the $H\,t$-integral of f exists for certain classes of functions which are not measurable on S. In this case the Lebesgue integral does not exist, nor does the Pierpont integral, since the existence of the latter insures the measurability[4] of f.

Let A and B be any two bounded sets, α and β sets of open intervals enclosing A and B respectively. If for every ϵ it is possible to find α and β such that $l\,(\alpha\,\beta) < \epsilon$ we shall say that the sets A and B are *separated.*

Let f be any function defined on the bounded set S and a any real number. Let S_a^f and S_f^a be the parts of S for which $f \geqq a$, $f < a$

[1] Received July 9, 1931.

[2] Proceedings of the London Mathematical Society, vol. 31, Part I, p. 18.

[3] Bull. Amer. Math. Soc., vol. 24, 1917, p. 128.

[4] Corollary III, Lemma II, below.

respectively. If for every a the sets S_a^f and S_f^a are separated we shall say that f is *separable relatively to* S.

Let f be bounded on the bounded set S. Let $s_1, s_2, \cdots$ be a finite or countably infinite sequence of separated sub-divisions of S. Let M_i and m_i be respectively the least upper bound and greatest lower bound of f on s_i. If for all possible such sub-divisions the greatest lower bound of $\sum M_i l s_i$ is equal to the least upper bound of $\sum m_i l s_i$, then this common value is the Pierpont integral of f over S. If the sets S, s_i, and the function f are measurable, then the common value of these two bounds is the Young integral of f over S.

3. Relations between Hildebrandt integrals and Pierpont integrals.

THEOREM I. *If f is bounded and separable relatively to the bounded set S, then both the Hildebrandt integral and the Pierpont integral of f over S exist and the two are equal.*

This is an immediate consequence of the following necessary and sufficient conditions for the existence of a Pierpont integral.

LEMMA I. *Let f be bounded on the bounded set S. Let s_{ni} $(i = 1, 2, \cdots)$ be any sequence of separated sub-divisions of S, M_{ni} and m_{ni} the least upper bound and greatest lower bound respectively of f on s_{ni}. Let*

$$\Sigma_n = \sum_i M_{ni} l s_{ni}, \qquad \sigma_n = \sum_i m_{ni} l s_{ni}.$$

It is then necessary and sufficient for the existence of the Pierpont integral of f over S that there be at least one sequence of sub-divisions for which

$$\lim_{n\to\infty} \Sigma_n = \lim_{n\to\infty} \sigma_n.$$

It is easily verified that Σ_n and σ_n exist for every n. We show that if s_{ni} and s'_{ni} are any two sequences of the type mentioned in the lemma, then every $\Sigma_n \geqq$ any σ'_n. Suppose this were not the case. Then there exists $d>0$, Σ_n, and $\sigma'_{n'}$ such that

$$\sigma'_{n'} - \Sigma_n > d.$$

Since f is bounded, it is easily verified that for a given Σ_n and $\sigma'_{n'}$ there exists for an arbitrary ε a value k of i and separated sub-divisions $s_1, s_2, \cdots, s_k$ and $s'_1, s'_2, \cdots, s'_k$ of S for which, if M_i and m_i are the greatest lower bound and least upper bound of f on s_i and s'_i respectively,

$$\Sigma_k = \sum_{i=1}^{k} M_i l s_i, \quad \text{and} \quad \sigma_k = \sum_{i=1}^{k} m_i l s'_i$$

differ respectively from Σ_n and $\sigma'_{n'}$ by not more than ε. We have then

$$\sigma_k - \Sigma_k > d - 2\varepsilon. \tag{3}$$

But it can be shown that any $\Sigma_k \geqq$ every σ_k. Let $e_{ij} = s_i s'_j$. Then, since all the sets involved are separated,

$$\Sigma_k = \sum_{i=1}^{k} M_i \sum_{j=1}^{k} l\, e_{ij}, \quad \text{and} \quad \sigma_k = \sum_{j=1}^{k} m_j \sum_{i=1}^{k} l\, e_{ij},$$

and it can be easily verified that $M_i\, l\, e_{ij} \geqq m_j\, l\, e_{ij}$. This shows that $\Sigma_k \geqq \sigma_k$ which, with (3), leads to a contradiction. We thus have obtained the desired result that every $\Sigma_n \geqq$ any σ'_n. If then for any sequence of separated sets s_{ni}, $\lim \Sigma_n = \lim \sigma_n$ it follows that this common limit is at the same time the greatest lower bound of $\sum M_i\, l\, s_i$ and the least upper bound of $\sum m_i\, l\, s_i$ for all possible finite or countably infinite separated sub-divisions s_i of S. Hence, by the definition given above, this common limit is the Pierpont integral of f over S.

The conditions of the lemma are also necessary. If the Pierpont integral of f over S exists it is evidently possible to find two sub-divisions of S into separated sets $s_1, \cdots, s_k$ and $s'_1, \cdots, s'_k$ for which

$$(1) \qquad \sum_{i=1}^{k} M_i\, l\, s_i - P\int_S f\,dx < \varepsilon, \quad \text{and} \quad P\int_S f\,dx - \sum_{i=1}^{k} m_i\, l\, s'_i < \varepsilon.$$

Set $e_{ij} = s_i s'_j$. Then, since all the sets involved are separated, we get from (1)

$$(2) \qquad \sum_{i=1}^{k} M_i \sum_{j=1}^{k} l\, e_{ij} - \sum_{j=1}^{k} m_j \sum_{i=1}^{k} l\, e_{ij} < 2\varepsilon.$$

If now the sets e_{ij} are ordered in the sequence $s_{\varepsilon 1}, s_{\varepsilon 2}, \cdots$, we get from (2), with a suitable change of notation and by taking into account the fact that M_i and m_j are an upper bound and a lower bound respectively of f on e_{ij}, that

$$\sum_{\tau=1}^{k^2} M_\tau\, l\, s_{\varepsilon\tau} - \sum_{\tau=1}^{k^2} m_\tau\, l\, s_{\varepsilon\tau} < 2\varepsilon.$$

Then, taking a sequence of values of ε tending to zero and $s_{\varepsilon_n \tau}$ the corresponding sequence of sub-divisions of S, we readily arrive at a sequence of sub-divisions satisfying the conditions of the lemma.

We note that if the sets S, s_{ni}, and the function f are all measurable, then Lemma I gives the necessary and sufficient conditions for the existence of the Young integral of f over S.

The next step in the proof of Theorem 3.1 consists in showing that if f is separable relatively to S then there exists at least one sequence of sub-divisions s_{ni} which satisfy the conditions of the foregoing lemma. Let $A_n = a_{n0}, a_{n1}, \cdots, a_{nn}$ be a sequence of consecutive[5] sub-divisions

[5] A_{n+1} is obtained by retaining the points of A_n and inserting additional points, but not more than one additional point between $a_{n,i-1}$ and a_{ni} $(i = 1, 2, \cdots, n)$.

of the range of f, and s_{ni} the part of S for which $a_{n,i-1} \leqq f < a_{ni}$. Clearly we have

$$\sum_{i=1}^{n} a_{n,i-1}\, l s_{ni} \leqq \sum_{i=1}^{n} m_{ni}\, l s_{ni} \leqq \sum_{i=1}^{n} M_{ni}\, l s_{ni} \leqq \sum_{i=1}^{n} a_{ni}\, l s_{ni}.$$

But as n increases the first and last members of this inequality approach each other, which shows that the sequence of sub-divisions s_{ni} has the desired properties. Consequently the Pierpont integral of f over S exists. The foregoing inequalities also show that

$$P\int_S f\,dx = \lim_{n\to\infty} \sum a_{ni}\, l s_{ni}.$$

But $\sum_{i=1}^{n} a_{ni}\, l s_{ni}$ decreases with n. Furthermore, it is possible to show by methods used elsewhere[6] that for a given ε, if $a_1, a_2, \cdots, a_m$ is a sub-division of the range of f with norm sufficiently small and s_i the part of S for which $a_{i-1} \leqq f < a_i$, then $\sum a_i\, l s_i$ does not exceed $\sum a_{ni}\, l s_{ni}$ by more than ε. It is also evident that

$$\sum_{i=1}^{n} a_i\, l s_i \geqq P\int_S f\,dx.$$

From these considerations it follows that the Hildebrandt integral of f over S exists, and that

$$Ht\int_S f\,dx = P\int_S f\,dx.$$

4. The Hildebrandt integral of functions separable relatively to S.

THEOREM II. *If f_1 and f_2 are functions which are bounded and separable relatively to the bounded set S, then f_1+f_2 and $f_1 f_2$ are separable relatively to S. So also is f_1/f_2 provided f_2 is different from zero.*

Let A_n be a sequence of consecutive sub-divisions covering the range of f_1, f_2 and f_1+f_2. Let s^{ν}_{ni} be the part of S for which $a_{n,i-1} \leqq f_{\nu} < a_{ni}$, and let $\varphi^{\nu}_n = a_{ni}$ on s^{ν}_i $(\nu = 1, 2)$. Then $\varphi_n = \varphi^1_n + \varphi^2_n$ is separable relatively to S. To show this let t be fixed arbitrarily small, and determine finite sets of non-overlapping cells ω^{ν}_{ni} containing points of s^{ν}_{ni} where ω^{ν}_{ni} and ω^{ν}_{nj} have no points in common for $i \neq j$, and where

$$|l\omega^{\nu}_{ni} - l s^{\nu}_{ni}| < t, \quad \text{and} \quad l(s^{\nu}_{ni} - \omega^{\nu}_{ni} s^{\nu}_{ni}) < t.$$

If $e = \sum_{i=1}^{n} (\omega^1_{ni} s^1_{ni} + \omega^2_{ni} s^2_{ni})$, then $l(s-e) < 2nt$. Let D be a bounded domain containing the set S and the cells ω^{ν}_{ni}. Set $\psi^{\nu}_n = a_{ni}$ on ω^{ν}_{ni}, and $\psi^{\nu}_n = 0$ elsewhere on D. Then

[6] Proceedings of the London Mathematical Society, loc. cit. p. 20.

$$\psi_n = \psi_n^1 + \psi_n^2 = \varphi_n = \varphi_n^1 + \varphi_n^2$$

on e. But since ψ_n is a step function on D, it is separable relatively to every set on D, and consequently separable relatively to e. Hence φ_n is separable relatively to this set. Furthermore, by taking t sufficiently small, le can be made arbitrarily near to lS, and from this it follows that φ_n is separable relatively to S. Let $f = f_1 + f_2$, and suppose that for some number a the sets S_a^f and S_f^a are not separated. From the set of consecutive sub-divisions A_n choose a monotone increasing sequence $a_{n1}^1, a_{n2}^2, \cdots$ which converges to a. Set $a_{n_j}^j = a_j$. Then $S_f^{a_j} \supset S_f^{a_{j-1}}$. Also $\lim S_f^{a_j} = S_f^a$. Hence, by a theorem of W. H. Young,[7] for j sufficiently large $lS_f^{a_j}$ is arbitrarily near to lS_f^a, and consequently $S_f^{a_j}$ and S_a^f are not separated. Hence if $(a - a_j)/2 = \eta$ we can choose n and k so that $\varphi_n - f < \eta$ on S, and at the same time $a - a_{nk} < \eta$. Then, since $\varphi_n \geqq f$, it is easily verified that $S_{\varphi_n}^{a_{nk}} \supset S_f^{a_j}$, and that $S_{a_{nk}}^{\varphi_n} \supset S_a^f$. Consequently φ_n is not separable relatively to S. But this is a contradiction. We conclude, therefore, that $f = f_1 + f_2$ is separable relatively to S. In a similar manner the same can be shown for $f_1 f_2$, and also for f_1/f_2 if $f_2 \neq 0$.

THEOREM III. *If the bounded functions f_1 and f_2 are separable relatively to the bounded set S, then the Hildebrandt integral of $f_1 + f_2$ over S exists, and*

$$Ht\int_S (f_1 + f_2)\, dx = Ht\int_S f_1\, dx + Ht\int_S f_2\, dx.$$

The truth of this theorem follows readily from the definitions involved.

THEOREM IV. *Let f be bounded and separable relatively to the bounded set S which is contained on the bounded domain D. About each point p of S as mid-point let ω_k be a sequence of cells with side length a_k where a_k tends to zero as k increases. Let $S\omega_k = s_k$. Then for all points p of S except a null set the ratio*

$$Ht\int_{s_k} f\, dx / ls_k$$

converges to $f(p)$.

First let f be greater than zero, and let G_λ be the part of S which is such that for each point p of G_λ there is at least one sequence of cells for which

$$Ht\int_{s_k} \frac{f\, dx}{l\, s_k} < f(p)\,(1 - \lambda) \qquad (k = 1, 2, \cdots). \tag{1}$$

Suppose $l\, G_\lambda > 0$. It is then easy to show that

$$Ht\int_{G_\lambda} f\, dx = q > 0.$$

[7] Theory of Sets of Points (Cambridge, 1906), Theorem 28, p. 104.

Now divide the range of f on S at the points $a_1, a_2, \cdots, a_n$ where for an arbitrary fixed η we have

$$a_i - a_{i-1} < \eta. \tag{2}$$

Let s_i be the part of S for which $a_{i-1} \leqq f < a_i$. Then s_i and s_j are separated for $i \neq j$. Let $g_i' = s_i G_\lambda$. About each point p_i' of g_i' there exists a sequence of cells ω_k with p_i' as mid-point for which (1) holds, where $s_k = S\omega_k$. But since s_i and s_j are separated, it is possible to choose, for an arbitrary ε_i, a part g_i of g_i' with $l g_i' - l g_i < \varepsilon_i$ which is such that, if $s_{ik} = s_i \omega_k$, we have, for p_i a point of g_i,

$$H t \int_{s_{ik}} \frac{f\,dx}{l\, s_{ik}} < f(p_i)(1-\lambda) \qquad (k = 1, 2, \cdots). \tag{3}$$

From the infinite set of cells containing g_i we can, by means of Vitali's theorem as stated by Hildebrandt[8], choose a finite set $\omega_1, \omega_2, \cdots$ such that

$$l g_i - \sum_n l g_i \omega_n < \varepsilon_i. \tag{4}$$

Then for each of this finite set we have, for p_{in} a point of g_i which is a mid-point of ω_n,

$$H_t \int_{s_{in}} \frac{f\,dx}{l\, s_{in}} < f(p_{in})(1-\lambda). \tag{5}$$

It is easily shown that there exists on s_{in} a point p_{in}' for which

$$f(p_{in}')\, l\, s_{in} \leqq H t \int_{s_{in}} f\,dx. \tag{6}$$

From (2) $f(p_{in}) = f(p_{in}') + t_{in}$, where $|t_{in}| < \eta$. Then (5) becomes, after dividing through by $f(p_{in})$,

$$H t \int_{s_{in}} \frac{f\,dx}{\{l\, s_{in} f(p_{in}') + s_{in}\, t_{in}\}} < 1 - \lambda. \tag{7}$$

If we carry out the above process for all values of i, sum the numerator of the right hand side of (7) with respect to i, and then with respect to n, and likewise sum the denominator, the resulting fraction is still not greater than $1-\lambda$. But by properly choosing ε_i the numerator can be made arbitrarily near to q; while from (6) and the fact that $|t_{in}| < \eta$ the denominator does not exceed the numerator by more than $\eta\, l\, S$, where η is at our disposal. Thus the supposition that $l G_\lambda > 0$ leads to a contradiction. We conclude, therefore, that $l G_\lambda = 0$. If in (1) we replace $1-\lambda$ by $1+\lambda$ and $<$ by $>$, the set G_λ for which the resulting inequality holds can, in a similar way, be shown to be a null set. Then taking a sequence of values of λ tending to zero we arrive at the desired result for $f > 0$.

[8] Bull. Amer. Math. Soc., vol. 32, no. 5, p. 447.

For f unrestricted as to sign it is only necessary to apply the above reasoning to $\varphi = f +$ constant > 0.

5. **Relations between sets.** Let A and B be any two bounded sets. Let each point of A be the mid-point of a sequence of cells ω_k of side length a_k where a_k tends to zero. Let $A_k = A\omega_k$, and $B_k = B\omega_k$. Let

$$\varrho_A(\omega_k, B) = \frac{lB_k}{lA_k}.$$

By interchanging the roles of the sets A and B in the foregoing, obtain $\varrho_B(\omega_k, A)$.

I. *At each point of A except a null set $\varrho_A(\omega_k, B)$ is defined for all values of k, and at each point of B except a null set $\varrho_B(\omega_k, A)$ is defined for all values of k.*

This follows from the fact that the outer metric density of a set is unity at each point of the set except a null set.[9]

Let A_B^0 be the part of A for which $\overline{\lim}\, \varrho_A(\omega_k, B) = 0$ and A_B^+ the part of A for which $\overline{\lim}\, \varrho_A(\omega_k, B) > 0$. Similarly define B_A^0 and B_A^+.

II. *Let E be a part of A_B^+ with $lE = \lambda > 0$. Then for a given $\varepsilon > 0$ there exists a number $d > 0$, a part E_d of E with $lE_d > \lambda - \varepsilon$, and for each point p of E_d a sequence of cells ω_k with p as mid-point for which a_k, the side length of ω_k, tends to zero, and for which*

$$\frac{lB_k}{lA_k} \geqq d \qquad (k = 1, 2, \cdots).$$

Let $d_1, d_2, \cdots$ be a sequence of positive numbers tending to zero monotonically. Let E_i be the part of E for which

$$\overline{\lim} \frac{lB_k}{lA_k} > d_i. \tag{1}$$

It is then evident that $E_i \supset E_{i-1}$, and $\lim E_i = E$. Consequently, by Young's theorem[10] $\lim lE_i = lE$. We can, therefore, take i sufficiently large to insure that $lE_i > \lambda - \varepsilon$. For such a value of i let $d_i = d$ and $E_i = E_d$. Then from (1) for each point of E_d we have

$$\overline{\lim} \frac{lB_k}{lA_k} > d. \tag{2}$$

Then from any sequence ω_k with a point E_d as center for which (2) holds it is evidently possible to choose a sub-sequence for which

[9] Hildebrandt, Bull. Amer. Math. Soc., vol. 32, no. 5, p. 451.

[10] Loc. cit.

$$\frac{lB_k}{lA_k} \geqq d \qquad (k = 1, 2 \cdots).$$

Evidently II is also valid when the roles of the sets A and B are interchanged.

III. *The sets A_B^0 and B are separated. So also are the sets B_A^0 and A.*

For points of A_B^0 $\lim \varrho_A(\omega_k, B) = 0$. Hence about each such point as mid-point there exists a sequence of cells ω_k such that

$$\frac{lB_k}{lA_k} < \varepsilon \qquad (k = 1, 3, \cdots). \tag{1}$$

It is now possible to apply Vitali's theorem to the set of cells enclosing A_B^0 for which (1) holds, and obtain a finite set $\omega_1, \omega_2, \cdots$ without common points and for which

$$lA_B^0 - \sum_n l\omega_n A_B^0 < \varepsilon. \tag{2}$$

But from (1)

$$\sum_n lB_n < \varepsilon \sum_n lA_n < \varepsilon lA,$$

from which the desired result immediately follows. In a similar way it is possible to prove that B_A^0 is separated from A.

IV. *There is no part of one of the sets A_B^+ and B_A^+, except null parts, which is separated from the other set.*

Suppose E is a part of A_B^+ with $lE = \lambda > 0$, which is separated from B. Then, by II, we can find a number $d > 0$, a part E_d of E with $lE_d > \lambda - \varepsilon$, and for each point p of E_d a sequence of cells ω_k such that

$$\frac{lB_k}{lA_k} \geqq d \qquad (k = 1, 2, \cdots). \tag{1}$$

Since by hypothesis E_d is separated from B it is possible to put E_d and B in open sets of cells α_d and β respectively in such a way that

$$l(\alpha_d, \beta) < \varepsilon. \tag{2}$$

Then $\alpha' = \alpha_d - \beta\alpha_d$ contains no points of B. But since $lE_d > \lambda - \varepsilon$ it follows from (2) that E', the part of E_d on α', is such that $lE' > \lambda - 2\varepsilon$. But for each point of E' (1) holds. Hence about each point of E' as mid-point we can find a sequence of cells ω_k on α' for which

$$\frac{lB_k}{lA_k} \geqq d \qquad (k = 1, 2, \cdots).$$

Applying Vitali's theorem to the set of cells thus obtained, we easily arrive at the fact that there are points of B on α'. Hence we conclude that

there is no part of A_B^+ which is separated from B. It then follows from III that no part of A_B^+ is separated from B_A^+. Similarly it can be proved that there is no part of B_A^+, except null parts, which is separated from A_B^+.

V. *The sets A_B^0 and A_B^+ are separated; so also are the sets B_A^0 and B_A^+.*

Suppose A_B^0 not separated from A_B^+. Let α_0 and α_+ be sets of open cells containing A_B^0 and A_B^+ respectively. Then, since these sets are not separated, the lower limit of $l(\alpha_0 \alpha_+) = \lambda > 0$, and from this it readily follows that for every α_0 and α_+ we have

$$(1) \qquad l(\alpha_0 \alpha_+ A_B^+) \geqq \lambda, \qquad l(\alpha_0 \alpha_+ A_B^0) \geqq \lambda.$$

The first inequality of (1), in conjunction with IV, shows that for every α_0 and α_+ we have $l(\alpha_0 \alpha_+ B_A^+) \geqq \lambda$, and this, in conjunction with the last inequality of (1), shows that A_B^0 and B_A^+ are not separated. We thus have a contradiction of III, and can conclude, therefore that A_B^0 is separated from A_B^+. Similarly it can be shown that B_A^0 is separated from B_A^+.

VI. $lA_B^+ = lB_A^+ = l(A_B^+ + B_A^+)$.

If we suppose that $lA_B^+ = lB_A^+ + c$, $c > 0$, it can then be shown by methods similar to the foregoing that there is a part E of A_B^+ which is separated from B_A^+. But this contradicts IV. Thus $lA_B^+ \leqq lB_A^+$. Similarly it can be shown that $lB_A^+ \leqq lA_B^+$. Hence $lA_B^+ = lB_A^+$. In order to complete the proof of VI we first note that $l(A_B^+ + B_A^+) \geqq lA_B^+$. A supposition that $l(A_B^+ + B_A^+) = lA_B^+ + c$, $c > 0$, readily leads to the fact that there is a part of B_A^+ which is separated from A_B^+. But this again contradicts IV, and the truth of VI is established.

VII. *Let λ_{AB} be the lower limit of $l(\alpha\beta)$ for all possible enclosures α and β of A and B respectively. Then $l(A_B^+ + B_A^+) = \lambda_{AB}$.*

If each point of A is in some cell of a set of open cells α, and each point of B is in some cell of a set of open cells β, then $\alpha - \alpha\beta$ contains no points of B, and, on account of IV, no points of A_B^+. Similarly $\beta - \alpha\beta$ contains no points of B_A^+. Thus $A_B^+ + B_A^+ \subset \alpha\beta$, and it readily follows from this that $l(A_B^+ + B_A^+) \leqq \lambda_{AB}$. Suppose that $l(A_B^+ + B_A^+) = \lambda = \lambda_{AB} - c$, $c > 0$. Put $A_B^+ + B_A^+$ in a set of open cells γ where $l\gamma < \lambda + c/2$. The set A_B^0 is separated from A_B^+ and from B, and the set B_A^0 is separated from B_A^+ and from A. Hence we can put A_B^0 in α_0 and B_A^0 in β_0 where $l(\alpha_0 \beta_0) < c/2$. We shall then have A in $\alpha = \gamma + \alpha_0$ and B in $\beta = \gamma + \beta_0$, and evidently $l(\alpha\beta) < \lambda + c < \lambda_{AB}$, which is a contradiction. The truth of VII then follows.

VIII.
$$l(A+B) = lA_B^0 + lB_A^0 + l(A_B^+ + B_A^+),$$
$$lA + lB = lA_B^0 + lB_A^0 + 2l(A_B^+ + B_A^+),$$
$$lA + lB = l(A+B) + l(A_B^+ + B_A^+).$$

The first and second of these relations are immediate consequences of III—VII, while the last is obtained by combining the first and second.

6. Functions not separable relatively to the set over which they are defined. We are now in a position to prove

THEOREM V. *Let f be bounded on the bounded set S, and let $S = S_1 + S_2 + \cdots + S_p$ where f is separable relatively to S_i $(i = 1, 2, \cdots, p)$. Then the Hildebrandt integral of f over S exists.*

We carry through the proof for $p = 2$. Except for more complicated notation there are no further difficulties when $p > 2$.

Let A_n be a sequence of consecutive sub-divisions of the range of f on S, and let $s_{n\nu i}$ be the part of S for which $a_{n,i-1} \leqq f < a_i$ $(\nu = 1, 2)$. Let ω be a cell with a point of $s_{n\nu i}$ as center. Since $s_{n\nu i}$ and $s_{n\nu j}$ are separated for $i \neq j$ it easily follows that the ratio

$$\frac{l\,\omega\, s_{n\nu j}}{l\,\omega\, s_{n\nu i}} = 0 \tag{1}$$

at every point of $s_{n\nu i}$ except a null set provided the side length of ω is sufficiently small. If then for all values of n and all values of i we remove from S the points for which the right side of (1) is not defined or is greater than zero for every ω, we alter S by at most a null set. In what follows we shall consider S so altered. Let

$$E_1(n) = \sum_{i=1}^{n} (s^+_{n1is_{n21}} + \cdots + s^+_{n1is_{n2,i-1}} + s^+_{n1is_{n2,i+1}} + \cdots + s^+_{n1is_{n2n}}),$$

$$E_2(n) = \sum_{i=1}^{n} (s^+_{n2is_{n11}} + \cdots + s^+_{n2is_{n1,i-1}} + s^+_{n2is_{n1,i+1}} + \cdots + s^+_{n2is_{n1n}}),$$

$$e_1(n) = s^+_{n11s_{n21}} + s^+_{n12s_{n22}} + \cdots + s^+_{n1ns_{n2n}},$$

$$e_2(n) = s^+_{n21s_{n11}} + s^+_{n22s_{n12}} + \cdots + s^+_{n2ns_{n1n}}.$$

It follows from IV of § 5 that there is no part of one of the sets $E_1(n)$ and $E_2(n)$, except null parts, which is separated from the other set. Likewise there is no part of one of the sets $e_1(n)$ and $e_2(n)$, except null parts, which is separated from the other set. While from the fact that f is separable relatively to S_1 and S_2, together with IV, it readily follows that $E_1(n) + E_2(n)$ is separated from $e_1(n) + e_2(n)$. We show next that (a) *The set $E_\nu(n+1)$ contains the set $E_\nu(n)$ $(\nu = 1, 2)$.*

Let p be a point of $E_1(n)$ and let ω_k be a sequence of cells with p as center for which a_k the side length of ω_k tends to zero. Let $s_{n\nu ik} = \omega_k\, s_{n\nu i}$ $(\nu = 1, 2)$. Then for some values of i and j, $i \neq j$, we have $p \subset s_{n1i}$, and

$$\overline{\lim}\, \frac{l\, s_{n2ik}}{l\, s_{n1ik}} \geqq d > 0. \tag{2}$$

But for some values of l and m, $l \neq m$

$$s_{n1ik} = s_{n+1,1lk} + s_{n+1,1,l+1,k},$$
$$s_{n2jk} = s_{n+1,2mk} + s_{n+1,2,m+1,k}.$$

We then have

$$\frac{ls_{n+1,2mk} + ls_{m+1,2,m+1,k}}{ls_{n+1,1lk} + ls_{n+1,1,l+1,k}} = \frac{ls_{n2jk}}{ls_{n1ik}}, \tag{3}$$

where either $p \subset s_{n+1,1lk}$ or $p \subset s_{n+1,l+1,k}$. Suppose the former to be true, and divide numerator and denominator of right hand side of (3) by $ls_{n+1,1lk}$. We thus get

$$\frac{\dfrac{ls_{n+1,2mk}}{ls_{n+1,1lk}} + \dfrac{ls_{n+1,2,m+1,k}}{ls_{n+1,1lk}}}{1 + \dfrac{ls_{n+1,1,l+1,k}}{ls_{n+1,1lk}}}.$$

It follows from (1) above that for k sufficiently large the denominator of this ratio is unity. Hence, on account of (2) and (3), either the upper limit of

$$\frac{ls_{n+1,2mk}}{ls_{n+1,1lk}} \text{ or of } \frac{ls_{n+1,2,m+1,k}}{ls_{n+1,1lk}}$$

is greater than or equal to $d/2$. But this makes p a point of $E_1(n+1)$.

(b) *The set* $e_\nu(n)$ *contains the set* $e_\nu(n+1)$ $(\nu = 1, 2)$.

This is evident from the definition of $e_\nu(n)$.

(c) *If* $E = \lim E_1(n) + E_2(n)$, *and* $e = \lim e_1(n) + e_2(n)$ *then* E *and* e *are separated.*

Suppose E' is a part of E which is not separated from e. For the sake of definiteness let $E_1 E' = E''$ where $lE'' > 0$. Since $E_1 = \lim E_1(n)$, we can use Young's theorem[11] to so fix n that $lE'' E_1(n) > 0$. Then, since $e_1(n) + e_2(n) \supset e$ it follows that $E_1(n)$ and $e_1(n) + e_2(n)$ are not separated. But this contradicts the fact that $E_1(n) + E_2(n)$ and $e_1(n) + e_2(n)$ are separated. We conclude, therefore, that E and e are separated.

(d) *If* E_1' *and* E_2' *are non separated parts of* E_1 *and* E_2, $E_\nu = ES_\nu$ $(\nu = 1, 2)$, *then* f *is not separable relatively to* $E' = E_1' + E_2'$.

Again making use of Young's theorem it is possible to fix n great enough to insure that $lE_1' E_1(n) > 0$, and that $lE_2' E_2(n) > 0$. Then $E_1' E_1(n)$ contains some part s_{ij} of $s^+_{n1is_{n2j}}$ $(i \neq j)$, and $E_2' E_2(n)$ contains some part s_{kl} of $s^+_{n2ks_{n1l}}$ $(k \neq l)$, where s_{ij} and s_{kl} are not separated. We note first that $k \neq i$. For if $k = i$, then $k \neq j$, and consequently s_{n2k} and s_{n2j} are not separated, which is a contradiction. Let e_i be the part

[11] Loc. cit.

of E for which $a_{n,i-1} \leqq f < a_{ni}$. Then $e_i \supset s_{ij}$ and $e_k \supset s_{kl}$. But since s_{ij} and s_{kl} are not separated it follows that e_i and e^k are not separated. Consequently f is not separable relatively to E'.

(e) *The Hildebrandt integral of f over $E = E_1 + E_2$ exists, and*

$$Ht\int_E f\,dx = Ht\int_{E_1} f\,dx + Ht\int_{E_2} f\,dx.$$

Let e_{ni} be the part of E for which $a_{n,i-1} \leqq f < a_{ni}$. In conformity with the notation of § 5, we set $e^0_{n1ie_{n2i}} = e^0_{n12i}$, and likewise define the corresponding sets e^+_{n12i}, e^0_{n21i}, and e^+_{n21i}. Let

$$H_1^n = \sum_{i=1}^{n} e^0_{n12i}, \quad H_2^n = \sum_{i=1}^{n} e^0_{n21i},$$

$$G_1^n = \sum_{i=1}^{n} e^+_{n12i}, \quad G_2^n = \sum_{i=1}^{n} e^+_{n21i}.$$

It can be easily verified by means of IV and V of § 3 that $H(n) = H_1^n + H_2^n$ is separated from $G(n) = G_1^n + G_2^n$. Also, the methods used in obtaining (a) can be used to show that $G_\nu^n \supset G_\nu^{n+1}$ $(\nu = 1, 2)$. Let G_1 and G_2 be the sets belonging to G_1^n and G_2^n respectively for all values of n. Then $G(n) \supset G = G_1 + G_2$. Let $g_i(n)$ be the part of $G(n)$ for which $a_{n,i-1} \leqq f < a_{ni}$. Then, since $g_i(n) = e^+_{n12i} + e^+_{n21i}$ it easily follows that $g_i(n)$ and $g_j(n)$ are separated for $i \neq j$. Let g_i be the part of G for which $a_{n,i-1} \leqq f < a_{ni}$. Hence, since $g_i(n)$ contains g_i, it follows that g_i and g_j are separated when $i \neq j$. But this easily leads to the conclusion that for any a_{ni} $G^f_{a_{ni}}$ and $G_f^{a_{ni}}$ are separated. It now follows that f is separable relatively to G. For if this were not the case there would exist some number a for which G^f_a and G_f^a were not separated. Proceeding as in the case of Lemma II, it is possible to choose from the set of sub-divisions A_n a monotone increasing sequence $a^1_{n_1}, a^2_{n_2}, \cdots$ which converges to a. Then, if $a^j_{n_j} = a_j$, by making use of the fact that $G^f_{a_j}$ and $G_f^{a_j}$ are separated and proceeding as in Lemma II, it is possible to contradict the supposition that G^f_a and G_f^a are not separated. Thus f is separable relatively to G. But if f is separable relatively to G it is evidently separable relatively to any part of G. It then follows from (d) that G_1 and G_2 are separated. The set H_2^n is separated from $G_1^n \supset G_1$, and therefore separated from G_1. Hence $H_2^n + G_2$ is separated from G_1, and since $\lim (H_2^n + G_2) = E_2$, it easily follows that G_1 is separated from E_2. But $G_1 \subset G_1^n$. And since for every n no part of G_1^n, except null parts, is separated from E_2, it follows that $l\,G_1 = 0$. In a similar manner it can be shown that $l\,G_2 = 0$. By methods indicated above it can be shown

that $H_\nu^n \supset H_{\nu-1}^n$ $(\nu = 1, 2)$, and from this and the facts that $\lim (H_\nu^n + G_\nu^n) = E_\nu$ and $l\, G_\nu = 0$, $(\nu = 1, 2)$, we are easily led to the conclusion that $\lim l\, H_\nu^n = l\, E_\nu$ $(\nu = 1, 2)$. Let $h_i(n')$ be the part of $H(n)$ for which $a_{n',i-1} \leqq f < a_{n'i}$ $(n' \geqq n)$. If $h_{\nu i}(n') = h_i(n')\, E_\nu$ $(\nu = 1, 2)$, it follows from the definitions of H_ν^n that $h_{1i}(n')$ and $h_{2i}(n')$ are separated. Consequently

$$\sum_{i=1}^{n'} a_{n'i}\, l\, h_i(n') = \sum_{i=1}^{n'} a_{n'i}\, l\, h_{1i}(n') + \sum_{i=1}^{n'} a_{n'i}\, l\, h_{2i}(n').$$

From this it easily follows that the Ht-integral of f over $H(n)$ exists, and

$$Ht\int_{H(n)} f\,dx = Ht\int_{H_1^n} f\,dx + Ht\int_{H_2^n} f\,dx.$$

The truth of (e) then follows from the fact that $l\, H_\nu^n$ tends to $l\, E_\nu$ $(\nu = 1, 2)$.

(f) *The function f is separable relatively to e, and if $e_\nu = e S_\nu$ $(\nu = 1, 2)$, then*

$$Ht\int_e f\,dx = Ht\int_{e_1} f\,dx = Ht\int_{e_2} f\,dx.$$

Let e_i be the part of e for which $a_{n,i-1} \leqq f < a_i$. It follows at once from the definitions of $e_1(n)$ and $e_2(n)$, and from the fact that $e_i(n) = e_{i1}(n) + e_{i2}(n)$ that e_i and e_j are separated when $i \neq j$. The above methods can now be used to establish the fact that f is separable relatively to e. The second part of f then follows from Theorem I, and from VI of § 5.

It is now possible to write

$$S = S_1 + S_2 = S_{12}^0 + S_{21}^0 + E + e, \tag{4}$$

where the four sets on the right are such that each is separated from the other three; where f is separable relatively to each of the sets S_{12}^0, S_{21}^0, and e; and where the Ht-integral of f over E exists. Hence we have

$$Ht\int_S f\,dx = Ht\int_{S_{12}^0} f\,dx + Ht\int_{S_{21}^0} f\,dx + Ht\int_E f\,dx + Ht\int_e f\,dx, \tag{5}$$

which establishes Theorem V.

COROLLARY. *Let f be defined and bounded on the bounded sets S_1 and S_2, and be separable relatively to each of these sets. It is then necessary and sufficient for*

$$Ht\int_{S_1+S_2} f\,dx = Ht\int_{S_1} f\,dx + Ht\int_{S_2} f\,dx$$

that $l\, e = 0$, or that $f = 0$ on e.

This follows at once from (e), (f), and (5).

LEMMA II. *Let f be bounded on the bounded set S. It is then necessary and sufficient for the existence of the Pierpont integral of f over S that f be separable relatively to S.*

The sufficiency of the condition follows from Theorem I. To show its necessity, let there be some number a for which S_a^f and S_f^a are not separated. Let $a_1, a_2, \cdots$ be a monotone increasing sequence converging to a. By the methods used above it is then possible to arrive at $E_1 = S_f^{a_i}$ such that, if $E_2 = S_a^f$, then E_1 and E_2 are not separated. Set $E = E_1 + E_2$ and let $e_1 + e_2 + \cdots + e_n$ be a separated subdivision of E for which $le_j > 0$ $(i = 1, 2, \cdots, n)$, M_i and m_i the least upper bound and greatest lower bound respectively of f on e_j. Let $\eta = a - a_i$. Then

$$\sum_{j=1}^{n} M_j le_j - \sum_{j=1}^{n} m_j le_j \geqq \eta \sum_{j=1}^{n} le_j.$$

It then follows from Lemma I that the Pierpont integral of f over S does not exist.

COROLLARY I. *Let f be bounded and separable relatively to the bounded sets S_1 and S_2. It is then necessary and sufficient for the existence of the Pierpont integral of f over S that $lE = 0$, where E is defined by* (4) *above.*

COROLLARY II. *Let f be bounded and separable relatively to the bounded sets S_1 and S_2. It is then necessary and sufficient for*

$$P\int_S f\,dx = P\int_{S_1} f\,dx + P\int_{S_2} f\,dx$$

that either $lE = le = 0$, or $lE = 0$ and $f = 0$ on e.

It is evident that separability of a function f relatively to a bounded domain D is equivalent to measurability of f on D. We thus get

COROLLARY III. *If the bounded function f defined on the bounded domain D is such that the Pierpont integral of f over D exists, then f is measurable on D.*

REMARK. If in Theorem V, S is a bounded domain D, and $S = S_1 + S_2$, where S_1 and S_2 are such that $lE > 0$, then f is not measurable on D, and consequently neither the Lebesgue integral nor the Young integral of f over D exist. Furthermore, it follows from Corollary I of Lemma II that the Pierpont integral of f over D does not exist. It is thus possible to have a function defined and bounded on the bounded domain D for which the Hildebrandt integral exists, but for which the same is not true of any of the other three integrals.

7. Functions which are not bounded on the bounded set S. Let f be defined on the bounded set S. Let $\cdots, a_{-1}, a_0, a_1, \cdots$ be a system of sub-divisions of the interval $(-\infty, \infty)$ for which $a_i - a_{i-1}$

tends to zero independently of i. Let s_i be the part of S for which $a_{i-1} \leqq f < a_i$. If there exists a number I such that

$$\left| I - \sum_{i=-\infty}^{i=\infty} a_i \, ls_i \right|$$

is arbitrarily small provided $a_i - a_{i-1}$ is sufficiently small, we shall say that this number I is the Ht-integral of f over S.

Let f be separable relatively to S, and let S_N be the part of S for which $|f| \leqq N$. If the Ht-integral of f over S exists it is easily verified that $Nl(S - S_N)$ tends to zero as N increases. Making use of this it is easy to prove that the necessary and sufficient conditions given above for

$$Ht\int_{S_1+S_2} f\,dx = Ht\int_{S_1} f\,dx + Ht\int_{S_2} f\,dx$$

also holds when the function f is unbounded but Ht-integrable. We further state the following theorem, the proof of which is easily obtained.

THEOREM VI. *Let $S = S_1 + S_2 + \cdots + S_p$, and let f be separable relatively to S_i $(i = 1, 2, \cdots, p)$. If there exists on S a positive function φ for which $|f| \leqq \varphi$ except for a null set, and if the Ht-integral of φ over S exists, then the Ht-integral of f over S exists.*

The principal theorems of integration have now been obtained for Hildebrandt integrals, and this without recourse to the theory of measure. Since sufficient has been accomplished to establish the general methods of procedure, we carry this phase of the discussion no further.

8. **Relations between sets in general and measurable sets, and between separable functions and measurable functions.** Let S be any set on the bounded domain D, and $\bar{S}$ its complement. If the sets S and $\bar{S}$ are separated we shall say that S is *measurable,* and shall call lS the measure of S. This is in conformity with the usual definition of measurability. We prove

I′. *If $E = S_{\bar{S}}^{+} + \bar{S}_{S}^{+}$ then E is measurable.*

Except for at most a null set, $\bar{E} = S_{\bar{S}}^{0} + \bar{S}_{S}^{0}$. Applying III and IV of § 5, together with the above definition of measurability, we obtain the desired result.

II′. $lD = l\bar{S}_{S}^{0} + lS$.

We have

$$D = \bar{S}_{S}^{0} + E + S_{\bar{S}}^{0} + e,$$

where e is a null set, and where the three remaining sets on the right are separated. Hence

$$lD = l\bar{S}_{S}^{0} + lE + lS_{\bar{S}}^{0}.$$

And from this, together with V and VI of § 5, we get

$$lD = l\bar{S}_S^0 + lS_S^+ + lS_S^0 = l\bar{S}_S^0 + lS.$$

In conformity with the usual definition of interior measure IS of a set S, we put $IS = lD - l\bar{S}$, and prove

III′. *The measure of E is the difference between the exterior and the interior measures of S.*

Making use of V and VI of § 3, we have

$$IS = lD - l\bar{S} = lD - l(\bar{S}_S^0 + \bar{S}_S^+) = lD - l\bar{S}_S^0 - l\bar{S}_S^+ = lD - l\bar{S}_S^0 - lE.$$

From this, together with II′, we get III′.

Let A and B be any two distinct bounded sets. Let $F = A_B^+ + B_A^+$, and let $E_{AB} = F_{\bar{F}}^0 + F_{\bar{F}}^+ + \bar{F}_F^+$. Then from V and I′ we have E_{AB} measurable, and from VI and VII we get

IV′. $lE_{AB} = \lambda_{AB}$.

The existence of a measurable set associated with the sets A and B the measure of which is equal to λ_{AB} has been pointed out by W. A. Wilson[12], but he gives no rules for determining such a set.

We now prove

V′. *Let S be a set contained on the bounded domain D, f a bounded function defined on S which is separable relatively to S. Then there exists a bounded function φ which is measurable on D, which is equal to f at each point of S except a null set, and for which*

$$L\int_D \varphi\, dx = Ht\int_S f\, dx.$$

Let the interval (l, L) be the range of f on S, A_n a consecutive sequence of sub-divisions of (l, L) for which $a_{n,i-1} - a_{ni} = (L-l)/2^{n-1}$, and s_{ni} the part of S for which $a_{n,i-1} \leqq f < a_{ni}$. For each s_{ni} determine the set e_{ni} corresponding to E of I′. Let e'_{ni} be the part of e_{ni} not on any e_{n1}, $e_{n2}, \cdots, e_{n,i-1}, e_{n,i+1}, \cdots, e_{nn}$. Set $\varphi_n = a_{ni}$ on e'_{ni}, and $\varphi_n = 0$ elsewhere on D. φ_n is then a monotonically decreasing sequence of measurable functions, and consequently converges to a measurable function φ. It is easily verified that φ fulfills the requirements of V′.

We note that Hildebrandt[13] has indicated the existence of a function such as φ, but gives no explicit scheme for its determination.

In conclusion we outline a method for determining a measurable function φ which is equal to f at every point of S except a null set, which method does not depend on the results of § 5 of this paper.

[12] Bull. Amer. Math. Soc., vol. 22, p. 384.

[13] Bull. Amer. Math. Soc., vol. 24, p. 129.

Let p be a point of a domain $D \supset S$ and ω a cell on D which contains p as an interior point. If for every such ω the set ωS is non measurable then we shall say that p is a point of non measurability[14] of S. If N is the totality of all such points it is easily verified that N is closed, and consequently measurable. If S contains no measurable component with measure greater than zero it is evident that $lN \geqq lS$. An example shows that $lN > lS$ can hold.

On the interval (0,1) let $\alpha = \alpha_1, \alpha_2, \cdots$ be an everywhere dense set of distinct open intervals for which $l\alpha < \varepsilon$. Then $\bar{\alpha}$ is non dense and closed, and $l\bar{\alpha} > 1 - \varepsilon$. Let S_i be a non measurable set on α_i and N the set of points of non measurability of $S = S_1 + S_2 + \cdots$. Evidently $lS < \varepsilon$, and since $N \supset \bar{\alpha}$, $lN > 1 - \varepsilon$. Also $IS = 1 - l\bar{S} \leqq 1 - l\bar{\alpha} < \varepsilon$.

To proceed with the determination of the function φ, let the range of f be sub-divided by A_n as in V'. Let N_{ni} be the points of non measurability of s_{ni}, and N'_{ni} the part of N_{ni} not contained on any $N_{n1}, N_{n2}, \cdots, N_{n,i-1}, N_{n,i+1}, \cdots, N_{nn}$. The function φ can then be determined by substituting the sets N'_{ni} for the sets e'_{ni} in the reasoning of V'.

[14] This formulation was suggested by D. C. Gillespie, who raised the question as to whether or not the measure of the set of all such points was necessarily equal to the difference between the outer measure and the inner measure of the set S. The example given below answers this question in the negative.

ACADIA UNIVERSITY, WOLFVILLE, NOVA SCOTIA.

ON THE SEMI-CONTINUITY OF DOUBLE INTEGRALS IN THE CALCULUS OF VARIATIONS.[1]

By E. J. McShane.[2]

In his *Fondamenti di Calcolo delle Variazioni* Tonelli has given a number of theorems on the semi-continuity properties of functionals of curves, establishing both necessary and sufficient conditions that integrals of the type usually considered in the calculus of variations be semi-continuous. Tonelli's work shows the importance of the rôle played in the calculus of variations by semi-continuity properties.

With regard to the semi-continuity properties of double integrals taken over a surface in three-space, certain necessary conditions have been established, both for the parametric form of the problem[3] and for the non-parametric form.[4] For the non-parametric problem a set of conditions sufficient for lower semi-continuity has been given by Tonelli.[5]

In the present paper we consider integrals of the form

$$F(S) \equiv \int\int f(x^1, x^2, x^3, X^1, X^2, X^3)\, du^1\, du^2,$$

where S is a rectifiable surface $x^i = x^i(u^1, u^2)$ $(i = 1, 2, 3)$, and the X^i are the Jacobians of the x^i with respect to the u's. Assuming that the integrand satisfies the usual conditions of continuity and homogeneity, we show that the integral is a functional of the surface only, and independent of the particular parameterization. We then find conditions on the integrand which are sufficient to insure the lower semi-continuity of $F(S)$.

1. **Generalities concerning continuous surfaces.** We shall in this paper be chiefly concerned with systems of three functions of two variables

$$x^1 = x^1(u^1, u^2), \quad x^2 = x^2(u^1, u^2), \quad x^3 = x^3(u^1, u^2). \tag{1.1}$$

[1] Received May 20, 1931.

[2] National Research Fellow.

[3] E. J. McShane, *On the necessary condition of Weierstrass in the multiple integral problem of the calculus of variations*, Annals of Mathematics, vol. 32 (1931), p. 723.

In this paper, as also in that of note [4], the condition obtained is stated to be necessary for a minimum, but is actually shown to be necessary for lower semi-continuity.

[4] E. J. McShane, *On the necessary condition of Weierstrass*, etc., Annals of Mathematics, vol. 32 (1931), p. 578.

[5] L. Tonelli, *Sur la semi-continuité des intégrales doubles du calcul des variations*, Acta Math. 53 (1929), p. 325.

For compactness such a system will henceforth be written in the form $x^i = x^i(u)$, or if desired simply $x = x(u)$; the use of the letter x will connote a triple of numbers, the use of u a pair of numbers. For example, the equation $u = u(\bar{u})$, or $u^i = u^i(\bar{u})$, will be understood to mean $u^1 = u^1(\bar{u}^1, \bar{u}^2)$, $u^2 = u^2(\bar{u}^1, \bar{u}^2)$. The Jacobians of the system (1.1) will be designated by capital letters:

$$(1.2)\quad X^1 \equiv \begin{vmatrix} \frac{\partial x^2}{\partial u^1} & \frac{\partial x^3}{\partial u^1} \\ \frac{\partial x^2}{\partial u^2} & \frac{\partial x^3}{\partial u^2} \end{vmatrix}, \quad X^2 \equiv \begin{vmatrix} \frac{\partial x^3}{\partial u^1} & \frac{\partial x^1}{\partial u^1} \\ \frac{\partial x^3}{\partial u^2} & \frac{\partial x^1}{\partial u^2} \end{vmatrix}, \quad X^3 \equiv \begin{vmatrix} \frac{\partial x^1}{\partial u^1} & \frac{\partial x^2}{\partial u^1} \\ \frac{\partial x^1}{\partial u^2} & \frac{\partial x^2}{\partial u^2} \end{vmatrix};$$

and if the x^i have a subscript, the X^i will bear the same subscript. A function of the type $\varphi(x^1, x^2, x^3, X^1, X^2, X^3)$ will be written $\varphi(x, X)$.

Double bars $\|\cdot\|$ will be used to denote distance; $\|u_1, u_2\|$ will mean $[(u_1^1-u_2^1)^2+(u_1^2-u_2^2)^2]^{1/2}$, $\|x_1, x_2\|$ will mean $[(x_1^1-x_2^1)^2+(x_1^2-x_2^2)^2+(x_1^3-x_2^3)^2]^{1/2}$, and for surfaces S_1, S_2 the symbol $\|S_1, S_2\|$ will mean the distance of S_1 from S_2 in a sense shortly to be defined. We shall also define

$$\|X\| \equiv [(X^1)^2+(X^2)^2+(X^3)^2]^{1/2}.$$

The nature of the symbol between the bars distinguishes unambiguously the particular sense of $\|\cdot\|$ each time it is used.

A system of three functions $x^i = x^i(u)$, defined, single-valued and continuous on a region B consisting of a Jordan curve and its interior, will be said to represent a continuous surface. (We do not assume that different values of u necessarily determine different values of x.) It is clear that there are infinitely many other systems $x^i = \bar{x}^i(\bar{u})$, $\bar{u}$ on $\bar{B}$, which represent what we would wish to call the same surface. To give a precise statement, we introduce a modification of the definition of distance of two surfaces given by Fréchet.[6] Let us assume that we have two representations of surfaces

$$(1.3)\qquad S:\ x^i = x^i(u),\quad u \text{ on } B,$$

$$(1.4)\qquad \bar{S}:\ x^i = \bar{x}^i(\bar{u}),\quad \bar{u} \text{ on } \bar{B}.$$

Suppose that T is a topological (i. e., one-to-one bicontinuous) mapping of $\bar{B}$ on B. Then if a point P traverses the boundary curve of B exactly once, its image $\bar{P}$ will traverse the boundary curve of $\bar{B}$ exactly once. If O be a point interior to B and $\bar{O}$ its image in $\bar{B}$, then the line

[6] M. Fréchet, *Sur la distance de deux surfaces*, Annales de la Soc. Polonaise de Math., 3, (1924), pp. 4–19.

segment OP will undergo a net rotation of $\pm 2\pi$; we suppose P constrained to move so that the rotation is $+2\pi$. Correspondingly the segment $\bar{O}\bar{P}$ will rotate through $\pm 2\pi$. If the rotation is $+2\pi$, we say that the mapping T is of *positive index*; if it is -2π, we say that T is of *negative index*.

Let us then suppose that T is a topological mapping of $\bar{B}$ on B, of positive index. If the points $P=(u)$ and $\bar{P}=(\bar{u})$ of B and $\bar{B}$ respectively correspond under T, then the expression $\|x(u), \bar{x}(\bar{u})\|$ has a definite maximum $M(T)$ as P ranges over B. The distance $\|S, \bar{S}\|$ is defined to be the greatest lower bound of $M(T)$ for all topological mappings T of positive index. We readily see that $\|S, \bar{S}\| = \|\bar{S}, S\| \geqq 0$, $\|S, S\| = 0$; and moreover if S_1, S_2, S_3 be any three continuous surfaces we have the triangle inequality $\|S_1, S_3\| \leqq \|S_1, S_2\| + \|S_2, S_3\|$.

We will say that (1.3) and (1.4) are *representations of the same surface* if and only if $\|S, \bar{S}\| = 0$. This will certainly be the case if there exists a topological mapping T of positive index such that for all corresponding points u and $\bar{u}$ we have $x^i(u) = \bar{x}^i(\bar{u})$. However, it is possible to have $\|S, \bar{S}\| = 0$ even if no such mapping exists.

In restricting ourselves to mappings of positive index, we have made a distinction analogous to distinguishing between the line segment AB and the segment BA. We do this because certain integrals of a type which we do not wish to exclude are invariant under changes of parameter of positive index, but not under changes of negative index. For example, let Q be the unit square $(0 \leqq u^1 \leqq 1,\ 0 \leqq u^2 \leqq 1)$. Define S and $\bar{S}$ by the relations

$$S\colon\ x^1 = 2u^1 - 1, \quad x^2 = u^2, \quad x^3 = 0 \qquad (u \text{ on } Q),$$
$$\bar{S}\colon\ x^1 = 1 - 2\bar{u}^1, \quad x^2 = \bar{u}^2, \quad x^3 = 0 \qquad (\bar{u} \text{ on } Q).$$

Under the definition of distance as originally given by Fréchet these surfaces are identical, under our definition they are distinct. We wish to consider them distinct, for if we define

$$F(S) \equiv \int (\|X\| + X^3)\, du,$$

we have $F(S) = 4$, $F(\bar{S}) = 0$.

We will say that a sequence $\{S_n\}$ of continuous surfaces approaches a continuous surface S_0 as a limit $(\lim S_n = S_0)$ if $\|S_n, S_0\|$ tends to zero. It follows that if $\lim S_n = S_0$, we can find for each S_n and for S_0 a representation S_j: $x^i = x_j^i(u)$, u on B $(j = 0, 1, 2, \cdots)$ such that $\lim x_n^i(u) = x_0^i(u)$ uniformly on B.

A continuous surface S will be called a *rectifiable surface* if among its representations there is one in the form (1.3) for which the functions $x^i(u)$

satisfy a Lipschitz condition. (The functions $x^i(u)$ are said to satisfy a Lipschitz condition if there exists a constant K such that the inequalities $|x^i(u_1)-x^i(u_2)| \leqq K\|u_1, u_2\|$ are satisfied for all points u_1, u_2 in B.) A representation of S satisfying these conditions is called a *typical representation* of S.

A continuous surface Π will be called a *polyhedron* if among its representations there is one in the form (1.3) for which the boundary of the region B is a simply closed polygon, and the region B can be subdivided into a finite number of triangles on each of which the functions $x^i(u)$ are linear in u^1 and u^2. Such a representation of Π will be called a *typical representation*. It is evident that every polyhedron Π is a rectifiable surface, and that its typical representations as a polyhedron serve also as typical representations as a rectifiable surface.

For every continuous surface S we define the area in the sense of Lebesgue as follows: Let Σ be a sequence $\{\Pi_n\}$ of polyhedra approaching S as limit. Each polyhedron Π_n has an area $A(\Pi_n)$ in the elementary sense, that is the sum of the areas of its triangular faces. For the sequence Σ we define $m(\Sigma)$ as $\liminf_{n\to\infty} A(\Pi_n)$. The area of S is then defined as the greatest lower bound of $m(\Sigma)$ for all possible sequences Σ of polyhedra approaching S. Since we are using a definition of distance different from that of Fréchet, this definition is different from that usually given, in which no account is taken of the indices of transformations involved. It is readily seen, however, that this definition is equivalent to the usual one.

Suppose that (1.3) is a typical representation of a rectifiable surface; then the six partial derivatives $\partial x^i/\partial u^j$ are all defined for almost all points of B, and are uniformly bounded and measurable. The same is therefore true of the Jacobians X^1, X^2, X^3, so that the integral

$$\int_B \|X\|\, du \equiv \int\int_B [(X^1)^2+(X^2)^2+(X^3)^2]^{1/2}\, du^1\, du^2$$

is well defined. It is known[7] that this integral is equal to the Lebesgue area of S; we shall however not make use of this fact, but shall in fact obtain a proof of it as a by-product of the methods here used.

2. **Generalities concerning integrals.** Suppose that we are given a function $\mathfrak{F}(x^1, x^2, x^3, p^1, q^1, p^2, q^2, p^3, q^3)$ which is continuous for all values of x^1, x^2, x^3 in a point set A of x-space and for all real values of the six remaining arguments. Then if the points of the rectifiable surface S, typically represented in the form S: $x^i = x^i(u)$, u on B, lie in A, the integral

[7] T. Radó, *Über das Flächenmaß rektifizierbarer Flächen*, Math. Annalen 100 (1928), pp. 445–479.

$$\int\int_B \mathfrak{F}\left(x^1(u),\ x^2(u),\ x^3(u),\ \frac{\partial x^1}{\partial u^1},\ \frac{\partial x^1}{\partial u^2},\ \frac{\partial x^2}{\partial u^1},\ \frac{\partial x^2}{\partial u^2},\ \frac{\partial x^3}{\partial u^1},\ \frac{\partial x^3}{\partial u^3}\right) du^1\, du^2$$

is well defined. If however we wish the value of such an integral to depend only on the surface and not on the particular representation of it, the function $\mathfrak{F}$ must be further restricted; it is in fact necessary[8] that $\mathfrak{F}$ depend only on the x^i and the Jacobians X^i, and moreover that it be positively homogeneous of degree 1 in the X^i:

$$\text{(2.1a)} \qquad \mathfrak{F}\left(x^1,\ x^2,\ x^3,\ \frac{\partial x^1}{\partial u^1},\ \cdots,\ \frac{\partial x^3}{\partial u^2}\right) \equiv f(x, X),$$

$$\text{(2.1b)} \quad f(x^1, x^2, x^3, kX^1, kX^2, kX^3) \equiv kf(x^1, x^2, x^3, X^1, X^2, X^3) \quad (k \geqq 0).$$

We shall henceforth assume that (2.1a) and (2.1b) are satisfied, and moreover that

(2.1c) $f(x, X)$, together with its first and second partial derivatives with respect to the X^i, is continuous for all x in A and all X such that $\|X\| \neq 0$.

Conditions (2.1) are sufficient[9] to assure that the integral shall not change in value under any change of parameters

$$\text{(2.2)} \qquad u^i = u^i(\overline{u})$$

satisfying a Lipschitz condition. However, we cannot immediately conclude from this that the integral is actually a function of the rectifiable surface alone, in the sense that its value is the same for all typical representations of such a surface; for two such representations may exist which are not related by any such transformation as (2.2). That such an integral is a function of the surface alone will be proved in Theorem I; in anticipation, we use the notation

$$F(S) \equiv \int_B f(x, X)\, du \equiv \int\int_B f(x^1, x^2, x^3, X^1, X^2, X^3)\, du^1\, du^2.$$

We shall make use of a fairly common modification of the tensor summation convention, and agree that the repetition of a *Greek-letter* subscript or superscript in any term will indicate the summation over all the values of that subscript or superscript; for example, $X_n^\alpha X_n^\alpha$ will mean $X_n^1 X_n^1 + X_n^2 X_n^2 + X_n^3 X_n^3$, the summation being extended to all the values of the α, but not to the n. With this convention, we define

$$E(x, X, \overline{X}) \equiv f(x, \overline{X}) - \overline{X}^\alpha f_{X^\alpha}(x, X).$$

[8] Vivanti, *Elementi del Calcolo delle Variazione*, pp. 153–161; or *Sull'equazione di Eulero*, etc., Rend. Cir. Mat. Palermo, vol. 33 (1912), p. 268.

[9] Loc. cit. [8]; the proof extends readily to Lipschitzian transformations.

As a consequence of the homogeneity relations (2.1) this can also be written

$$E(x, X, \bar{X}) = f(x, \bar{X}) - f(x, X) - (\bar{X}^\alpha - X^\alpha) f_{X^\alpha}(x, X). \tag{2.3}$$

This is well-defined for all x in A, all $\bar{X}$, and all X such that $\|X\| \neq 0$; it is positively homogeneous of degree 1 in $\bar{X}$, and of degree 0 in X.

The functional $F(S)$ will be called *positive quasi-regular* if $E(x, X, \bar{X}) \geqq 0$ for all x in A, all $\bar{X}$, and all X such that $\|X\| \neq 0$. It will be called *positive definite* if $f(x, X) > 0$ for every x in A and every X such that $\|X\| \neq 0$; *positive semi-definite* if $f(x, X) \geqq 0$ for every x in A and every X.

3. **Simplification of the problem.** We shall now turn our attention to sequences of rectifiable surfaces approaching a rectifiable surface as a limit. Stated in more detail, we shall consider systems of the form

$$S_0: \quad x^i = x_0^i(u), \qquad u \text{ on } B_0, \tag{3.1a}$$

$$S_n: \quad x^i = x_n^i(u), \qquad u \text{ on } B_n, \tag{3.1b}$$

in which each function satisfies some Lipschitz condition, and in which each region B_n can be mapped on B_0 by a topological mapping T_n of positive index such that for all points u and $\bar{u}$ corresponding under T_n

$$\|x_0(u), x_n(\bar{u})\| \leqq \eta_n, \quad \text{where} \quad \lim \eta_n = 0. \tag{3.1c}$$

Our principal object in this paper is to prove that if the function $f(x, X)$ satisfy certain conditions, then

$$\liminf \int_{B_n} f(x_n, X_n)\, du \geqq \int_{B_0} f(x_0, X_0)\, du. \tag{3.2}$$

In this section we shall show that if the function $f(x, X)$ satisfy the condition

$$f(x, X) \geqq 0 \tag{3.3}$$

for all x in a neighborhood of the points of S_0 and for all X, then in order to prove that inequality (3.2) holds for systems (3.1) we have only to show that (3.2) holds for certain systems of a simpler type than (3.1). We observe that under condition (3.3) we may assume that $f(x_n, X_n) \geqq 0$ for every n, for all the surfaces S_n save at most a finite number will lie in the neighborhood of S_0 in which (3.3) is satisfied.

We now effect four simplifications.

First. We can in (3.1) *restrict* B_0 *to be the unit square* Q: $0 \leqq u^1 \leqq 1$, $0 \leqq u^2 \leqq 1$; that is, we can show that if there exists a system (3.1) for which

$$\liminf \int_{B_n} f(x_n, X_n)\, du - \int_{B_0} f(x_0, X_0)\, du < 0, \tag{3.4}$$

$$\iint_B \mathfrak{F}\left(x^1(u),\ x^2(u),\ x^3(u),\ \frac{\partial x^1}{\partial u^1},\ \frac{\partial x^1}{\partial u^2},\ \frac{\partial x^2}{\partial u^1},\ \frac{\partial x^2}{\partial u^2},\ \frac{\partial x^3}{\partial u^1},\ \frac{\partial x^3}{\partial u^3}\right) du^1\, du^2$$

is well defined. If however we wish the value of such an integral to depend only on the surface and not on the particular representation of it, the function $\mathfrak{F}$ must be further restricted; it is in fact necessary[8] that $\mathfrak{F}$ depend only on the x^i and the Jacobians X^i, and moreover that it be positively homogeneous of degree 1 in the X^i:

$$(2.1\,a) \qquad \mathfrak{F}\left(x^1,\ x^2,\ x^3,\ \frac{\partial x^1}{\partial u^1},\ \cdots,\ \frac{\partial x^3}{\partial u^2}\right) \equiv f(x, X),$$

$$(2.1\,b) \quad f(x^1, x^2, x^3, kX^1, kX^2, kX^3) \equiv kf(x^1, x^2, x^3, X^1, X^2, X^3) \quad (k \geqq 0).$$

We shall henceforth assume that (2.1 a) and (2.1 b) are satisfied, and moreover that

(2.1 c) $f(x, X)$, together with its first and second partial derivatives with respect to the X^i, is continuous for all x in A and all X such that $\|X\| \neq 0$.

Conditions (2.1) are sufficient[9] to assure that the integral shall not change in value under any change of parameters

$$(2.2) \qquad u^i = u^i(\overline{u})$$

satisfying a Lipschitz condition. However, we cannot immediately conclude from this that the integral is actually a function of the rectifiable surface alone, in the sense that its value is the same for all typical representations of such a surface; for two such representations may exist which are not related by any such transformation as (2.2). That such an integral is a function of the surface alone will be proved in Theorem I; in anticipation, we use the notation

$$F(S) \equiv \int_B f(x, X)\, du \equiv \iint_B f(x^1, x^2, x^3, X^1, X^2, X^3)\, du^1\, du^2.$$

We shall make use of a fairly common modification of the tensor summation convention, and agree that the repetition of a *Greek-letter* subscript or superscript in any term will indicate the summation over all the values of that subscript or superscript; for example, $X_n^\alpha X_n^\alpha$ will mean $X_n^1 X_n^1 + X_n^2 X_n^2 + X_n^3 X_n^3$, the summation being extended to all the values of the α, but not to the n. With this convention, we define

$$E(x, X, \overline{X}) \equiv f(x, \overline{X}) - \overline{X}^\alpha f_{X^\alpha}(x, X).$$

[8] Vivanti, *Elementi del Calcolo delle Variazione*, pp. 153–161; or *Sull'equazione di Eulero*, etc., Rend. Cir. Mat. Palermo, vol. 33 (1912), p. 268.

[9] Loc. cit. [8]; the proof extends readily to Lipschitzian transformations.

As a consequence of the homogeneity relations (2.1) this can also be written

$$E(x, X, \bar{X}) = f(x, \bar{X}) - f(x, X) - (\bar{X}^\alpha - X^\alpha) f_{X^\alpha}(x, X). \tag{2.3}$$

This is well-defined for all x in A, all $\bar{X}$, and all X such that $\|X\| \neq 0$; it is positively homogeneous of degree 1 in $\bar{X}$, and of degree 0 in X.

The functional $F(S)$ will be called *positive quasi-regular* if $E(x, X, \bar{X}) \geqq 0$ for all x in A, all $\bar{X}$, and all X such that $\|X\| \neq 0$. It will be called *positive definite* if $f(x, X) > 0$ for every x in A and every X such that $\|X\| \neq 0$; *positive semi-definite* if $f(x, X) \geqq 0$ for every x in A and every X.

3. **Simplification of the problem.** We shall now turn our attention to sequences of rectifiable surfaces approaching a rectifiable surface as a limit. Stated in more detail, we shall consider systems of the form

$$S_0: \quad x^i = x_0^i(u), \qquad u \text{ on } B_0, \tag{3.1a}$$

$$S_n: \quad x^i = x_n^i(u), \qquad u \text{ on } B_n, \tag{3.1b}$$

in which each function satisfies some Lipschitz condition, and in which each region B_n can be mapped on B_0 by a topological mapping T_n of positive index such that for all points u and $\bar{u}$ corresponding under T_n

$$\|x_0(u), x_n(\bar{u})\| \leqq \eta_n, \quad \text{where} \quad \lim \eta_n = 0. \tag{3.1c}$$

Our principal object in this paper is to prove that if the function $f(x, X)$ satisfy certain conditions, then

$$\liminf \int_{B_n} f(x_n, X_n)\, du \geqq \int_{B_0} f(x_0, X_0)\, du. \tag{3.2}$$

In this section we shall show that if the function $f(x, X)$ satisfy the condition

$$f(x, X) \geqq 0 \tag{3.3}$$

for all x in a neighborhood of the points of S_0 and for all X, then in order to prove that inequality (3.2) holds for systems (3.1) we have only to show that (3.2) holds for certain systems of a simpler type than (3.1). We observe that under condition (3.3) we may assume that $f(x_n, X_n) \geqq 0$ for every n, for all the surfaces S_n save at most a finite number will lie in the neighborhood of S_0 in which (3.3) is satisfied.

We now effect four simplifications.

First. We can in (3.1) *restrict* B_0 *to be the unit square* Q: $0 \leqq u^1 \leqq 1$, $0 \leqq u^2 \leqq 1$; that is, we can show that if there exists a system (3.1) for which

$$\liminf \int_{B_n} f(x_n, X_n)\, du - \int_{B_0} f(x_0, X_0)\, du < 0, \tag{3.4}$$

then there exists a system (3.1) with $B_0 = Q$ for which inequality (3.4) holds.

Designate by -2ε the left member of inequality (3.4). We can construct in B_0 a finite set of non-overlapping squares $q_1, \cdots, q_m$ such that

$$\left| \int_{B_0 - \Sigma q_j} f(x_0, X_0)\, du \right| \leqq \varepsilon.$$

The mapping T_n maps $q_1, \cdots, q_m$ on a set of regions in B_n, which we call $q_1^n, \cdots, q_m^n$. Then

$$\begin{aligned} \liminf \sum_{j=1}^{m} \left[\int_{q_j^n} f(x_n, X_n)\, du - \int_{q_j} f(x_0, X_0)\, du \right] \\ \leqq \liminf \left[\int_{B_n} f(x_n, X_n)\, du - \int_{\Sigma q_j} f(x_0, X_0)\, du \right] \\ \leqq \liminf \left[\int_{B_n} f(x_n, X_n)\, du - \int_{B_0} f(x_0, X_0)\, du \right] + \varepsilon \leqq -\varepsilon, \end{aligned}$$

so that on at least one q_j we must have

$$\liminf \int_{q_j^n} f(x_n, X_n)\, du < \int_{q_j} f(x_0, X_0)\, du, \tag{3.5}$$

The square q_j can be mapped on the unit square Q by a linear mapping $w^i = u^i(\bar{u})$. Define $\bar{x}_0^i(\bar{u}) \equiv x_0^i(u(\bar{u}))$, $\bar{u}$ on Q; then

$$\int_Q f(\bar{x}_0, \bar{X}_0)\, d\bar{u} = \int_{q_j} f(x_0, X_0)\, du,$$

which with (3.5) gives us the desired result.

REMARK 1. By the homogeneity of the E-function we have for every set of numbers X^* the identity $E(x_0(u), X_0(u), X^*) = E(\bar{x}_0(\bar{u}), \bar{X}_0(\bar{u}), X^*)$.

Second. We can in (3.1) *restrict* B_n *to be the square* Q, *for each value of* n. For, let $\{\varepsilon_n\}$ be a sequence of positive numbers tending to zero. We map B_n conformally on the unit circle K_1: $\|u\| \leqq 1$. Then the functions

$$S_n\colon\ x^i = \bar{x}_n^i(\bar{u}), \quad \bar{u} \text{ on } K_1 \tag{3.6}$$

are uniformly continuous; hence we can find a $\delta_n > 0$ such that for each pair of points $\bar{u}_1$, $\bar{u}_2$ for which $\|\bar{u}_1, \bar{u}_2\| \leqq \delta_n$ we have

$$\|\bar{x}_n(\bar{u}_1), \bar{x}_n(\bar{u}_2)\| \leqq \varepsilon_n. \tag{3.7}$$

Moreover, if for each $r < 1$ we let S_n^r be that portion of S_n defined by (3.6) with u on the circle K_r: $\|u\| \leqq r$ and let B_n^r be the image in B_n of K_r, we find that the mapping of B_n^r on K_r is analytic, hence Lipschitzian,

so that the functions $x^i = \bar{x}_n^i(\bar{u})$, $\bar{u}$ on K_r, form a typical representation of S_n^r, and we have

$$\int_{K_r} f(\bar{x}_n, \bar{X}_n)\, d\bar{u} = \int_{B_n^r} f(x_n, X_n)\, du \leqq \int_{B_n} f(x_n, X_n)\, du. \tag{3.8}$$

Fixing r at the value $r(n) = 1 - \delta_n$, we map each point (ϱ, ϑ) of K_1 on the point $(r\varrho, \vartheta)$ of K_r; by (3.7) it follows that $\| S_n^{r(n)}, S_n \| \leqq \varepsilon_n$. Hence

$$0 \leqq \lim \| S_n^{r(n)}, S_0 \| \leqq \lim \| S_n, S_0 \| + \lim \varepsilon_n = 0,$$

so that the surfaces $S_n^{r(n)}$ approach S_0; and by (3.8)

$$\liminf \int_{K_r} f(\bar{x}_n(u), \bar{X}_n(u))\, du \leqq \liminf \int_{B_n} f(x_n(u), X_n(u))\, du.$$

Since K_r can readily be mapped on Q by a Lipschitzian transformation, we can represent $S_n^{r(n)}$ typically on Q.

Third. We can in (3.1) *restrict the surfaces* S_n *to be polyhedra.* We assume that each S_n is typically represented on Q, and select a sequence $\{\varepsilon_n\}$ of positive numbers tending to zero. Fixing upon a particular n, we successively subdivide the square Q into $4, 9, 16, \cdots, h^2, \cdots$ equal squares and divide each square into two triangles by a line from lower right to upper left corner. For each h we define

$$\Pi_{n,h}\colon\ x^i = x_{n,h}^i(u), \quad u \text{ on } Q,$$

by setting $x_{n,h}^i(u) = x_n^i(u)$ at each vertex, and linear in the interior of each of the $2h^2$ triangles. $\Pi_{n,h}$ is then inscribed in S_n. As h tends to ∞, the Jacobians $X_{n,h}^i$ converge[10] in measure to X_n^i, and are uniformly bounded; hence

$$\lim_{h\to\infty} \int_Q f(x_{n,h}(u), X_{n,h}(u))\, du = \int_Q f(x_n(u), X_n(u))\, du.$$

Also $\| \Pi_{n,h}, S_n \|$ tends to zero; so that we can find an $h(n)$ for which

$$\int_Q f(x_{n,h(n)}(u), X_{n,h(n)}(u))\, du \leqq \int_Q f(x_n(u), X_n(u))\, du + \varepsilon_n,$$
$$\| \Pi_{n,h(n)}, S_n \| \leqq \varepsilon_n;$$

whence

$$\liminf \int_Q f(x_{n,h(n)}(u), X_{n,h(n)}(u))\, du \leqq \liminf \int_Q f(x_n(u), X_n(u))\, du,$$
$$\lim \| \Pi_{n,h(n)}, S_0 \| = 0.$$

REMARK 2. We have in the second and third simplifications made no essential use of the assumption $f(x, X) \geqq 0$; for by selecting r large enough we could make

[10] For a proof of this see, e. g., H. Rademacher, *Über partielle und totale Differenzierbarkeit von Funktionen mehrerer Variablen*, Math. Annalen 81 (1920), pp. 52–63.

$$\left| \int_{K_r} f(\bar{x}_n, \bar{X}_n)\, du - \int_{B_n} f(x_n, X_n)\, du \right| \leqq \varepsilon_n,$$

which serves as well as (3.8) in the demonstration. Hence we have except for notation proved that if $f(x, X)$ satisfy conditions (2.1), then for every rectifiable surface S: $x^i = x^i(u)$, u on B, there exists a sequence $\{\Pi_n\}$ of polyhedra Π_n: $x^i = x_n^i(u)$, u on Q, such that

$$\lim \Pi_n = S, \qquad \lim \int_Q f(x_n, X_n)\, du = \int_B f(x, X)\, du.$$

Fourth. Let us say that a function $\varphi(u)$ defined on a polygonal region B is *quasi-linear* if it is continuous on B, and B can be subdivided into a finite number of triangles on each of which $\varphi(u)$ is linear. A mapping T: $u^i = u^i(\bar{u})$ of a region $\bar{B}$ on a region B will be called quasi-linear if $u^1(\bar{u})$ and $u^2(\bar{u})$ are each quasi-linear on $\bar{B}$. Our final simplification of the problem consists in showing that *the topological mappings T_n of the system* (3.1) (which are now mappings of Q on itself) *can be taken to be quasi-linear.*

Let $\{\varepsilon_n\}$ be a sequence of positive numbers tending to zero. Since S_0 is continuous on Q, we can find a $\delta_n > 0$ such that if u_1 and u_2 are points of Q for which $\|u_1, u_2\| \leqq \delta_n$, then

$$\|x_0(u_1), x_0(u_2)\| \leqq \varepsilon_n. \tag{3.9}$$

Let the mapping T_n be defined by the equations $u^i = u_n^i(\bar{u})$, and let T_n^*: $u = u_n^*(\bar{u})$ be a quasi-linear topological mapping of Q on itself such that $\|u_n(\bar{u}), u_n^*(\bar{u})\| \leqq \delta_n$. That such mappings exist follows from a lemma established by Franklin and Wiener.[11] We then have by (3.1) and (3.9)

$$\|x_0(u_n^*(\bar{u})), x_n(\bar{u})\| \leqq \eta_n + \varepsilon_n;$$

and since $\lim(\eta_n + \varepsilon_n) = 0$, the quasi-linear mapping T_n^* can be used instead of T_n.

Now if S_n be taken (as we have already seen possible) to be a polyhedron typically represented in the form $x^i = \bar{x}_n^i(\bar{u})$, $\bar{u}$ on Q, and the transformations T_n: $u^i = u_n^i(\bar{u})$ (and consequently their inverses $\bar{u}^i = \bar{u}_n^i(u)$) are quasi-linear, then the representation of S_n: $x^i = x_n^i(u) \equiv \bar{x}_n^i(\bar{u}_n(u))$ is also typical; and since T_n is Lipschitzian we have

$$\int_Q f(x_n, X_n)\, du = \int_Q f(\bar{x}_n, \bar{X}_n)\, d\bar{u}.$$

[11] P. Franklin and N. Wiener, *Analytical Approximations to Topological Transformations*, Trans. Am. Math. Soc., vol. 28 (1926), p. 764.

Hence, recalling (3.1c), *we can if we wish consider the* S_n *to be polyhedra typically represented in the form* S_n: $x^i = x_n^i(u)$, u *on* Q, *with the condition* $\lim x_n^i(u) = x_0^i(u)$ *uniformly on* Q.

4. **Convergent sequences of polyhedra.** We now proceed to establish

LEMMA 1. *If*

a) *the functions* $x_0(u)$ *satisfy a Lipschitz condition on a square* q;

b) *the functions* $x_n(u)$ *are quasi-linear on* q;

c) $\lim x_n^i(u) = x_0^i(u)$ *uniformly on* q;

then for every set of constants a_1, a_2, a_3 *we can find a sequence* $\{V_n\}$ *of subsets of* q *such that*

$$\lim \int_{V_n} a_\alpha X_n^\alpha \, du = \int_q a_\alpha X_0^\alpha \, du.$$

There is clearly no loss of generality in assuming $a_\alpha a_\alpha = 1$. Let b_i, c_i be six numbers such that the matrix

$$\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}$$

is orthogonal, and apply the transformation

$$\xi^1 = a_\alpha x^\alpha, \quad \xi^2 = b_\alpha x^\alpha, \quad \xi^3 = c_\alpha x^\alpha; \tag{4.1}$$

by direct calculation we find that for every set of Lipschitzian functions $x^i(u)$ we have $\Xi^1 = a_\alpha X^\alpha$ holding almost everywhere. Hence we need only find sets V_n such that

$$\lim \int_{V_n} \Xi_n^1 \, du = \int_q \Xi_0^1 \, du.$$

We shall retain the symbols x^i, X^i instead of using ξ^i, Ξ^i.

The transformation $x^i = x_0^i(u)$ $(i = 2, 3)$ carries the boundary of q into a rectifiable curve, which we call Γ_0, in the $x^2 x^3$ plane; likewise the transformations $x^i = x_n^i(u)$ $(i = 2, 3)$ carry the boundary of q into polygons Γ_n in the $x^2 x^3$ plane, and $\lim \Gamma_n = \Gamma_0$. Let (x^2, x^3) be a point not lying on Γ_0; its index[12] $\mu_0(x^2, x^3)$ with respect to Γ_0 is well defined, and[13]

$$\int_q X_0^1 \, du = \int \mu_0(x^2, x^3) \, dx^2 \, dx^3,$$

[12] *Ordnung* des Punktes; see, e. g., Osgood, *Lehrbuch der Funktionentheorie*, 5th ed., vol. I, p. 175.

[13] J. Schauder, *Über stetige Abbildungen*, Fundamenta Math., 12 (1928), pp. 47–74; in particular, Satz VIII.

the second integral being taken over the whole x^2x^3 plane. We define $\mu_n(x^2, x^3)$ analogously. The set of points of the x^2x^3 plane for which $\mu_0(x^2, x^3)$ is defined and not zero is a bounded open set, containing none of the points of Γ_0; so for every $\varepsilon > 0$ we can construct a finite number of squares $q_1, \cdots, q_h$ lying in that open set and such that

$$\left| \sum_{j=1}^{h} \int_{q_j} \mu_0(x^2, x^3)\, dx^2\, dx^3 - \int \mu_0(x^2, x^3)\, dx^2\, dx^3 \right| < \varepsilon. \tag{4.2}$$

There is a positive lower bound γ to the distance between points (x^2, x^3) lying in the set $\sum q_j$ and points (x^2, x^3) lying on Γ_0; and since $\lim \Gamma_n = \Gamma_0$, for all n greater than a certain n_0 we will have $\|\Gamma_n, \Gamma_0\| \leqq \gamma/3$. Then for every $n > n_0$ and every point (x^2, x^3) lying in $\sum q_j$ we have $\mu_n(x^2, x^3) = \mu_0(x^2, x^3)$. Fix n at any value $> n_0$. The edges of the flat polyhedron $x^1 = 0$, $x^2 = x_n^2(u)$, $x^3 = x_n^3(u)$ subdivide each square q_j into a finite number of polyhedral regions, which we suppose further subdivided into a finite number of non-overlapping triangles $T_1, \cdots, T_s$; $\sum T_j = \sum q_j$. Let (x^2, x^3) be a point interior of T_j, of index $\mu_n(x^2, x^3) = \mu_0(x^2, x^3)$. Then $\mu_n(x^2, x^3)$ is constant throughout T_j, since we can join any two points of T_j by a line-segment which does not intersect Γ_n. The triangle T_j is the image under the transformation $x^2 = x_n^2(u)$, $x^3 = x_n^3(u)$ of a point set in the u-plane consisting[14] of a finite number of triangles $\tau_{j,1}, \tau_{j,2}, \cdots, \tau_{j,m(j)}$; and

$$\sum_{k=1}^{m(j)} \int_{\tau_{j,k}} X_n^1\, du = \int_{T_j} \mu_n(x^2, x^3)\, dx^2\, dx^3.$$

No two triangles $\tau_{j,k}$ overlap; hence

$$\sum_{j=1}^{s} \sum_{k=1}^{m(j)} \int_{\tau_{j,k}} X_n^1\, du = \sum_{j=1}^{s} \int_{T_j} \mu_n(x^2, x^3)\, dx^2\, dx^3$$

$$= \sum_{j=1}^{h} \int_{q_j} \mu_n(x^2, x^3)\, dx^2\, dx^3 = \sum_{j=1}^{h} \int_{q_j} \mu_0(x^2, x^3)\, dx^2\, dx^3;$$

which by (4.2) differs from

$$\int \mu_0(x^2, x^3)\, dx^2\, dx^3 = \int_q X_0^1\, du$$

by less than ε.

Consider now a sequence $\{\varepsilon_j\}$ of positive numbers tending to zero. To each ε_j there corresponds an integer n_j such that for every $n > n_j$ we can define a set V_n^j (namely the set $\sum_j \sum_k \tau_{j,k}$ defined just above) such that

$$\left| \int_{V_n^j} X_n^1\, du - \int_q X_0^1\, du \right| \leqq \varepsilon_j.$$

[14] Loc. cit. [13].

We can assume the numbers $n_1, n_2, \cdots, n_j, \cdots$ to be monotonic increasing. Then for each j we define $V_n = V_n^j$ $(n_j < n \leqq n_{j+1})$. The sets V_n thus defined have the desired property.

REMARK. Suppose that after the rotation (4.1) we find $\int_q X_0^1 \, du > 0$; we can then select the V_n in such a way that at every point in V_n, $X_n^1 > 0$. For let n be large enough so that

$$\int_{V_n} X_n^1 \, du > 0,$$

and denote by V_n^+ that subset (consisting of a finite number of triangles) on which $X_n^1 > 0$; then

$$\int_{V_n^+} X_n^1 \, du \geqq \int_{V_n} X_n^1 \, du.$$

Now if we shrink each triangle of V_n^+ on itself continuously to zero, we must at some stage pass through a set V_n^* such that

$$\int_{V_n^*} X_n^1 \, du = \int_{V_n} X_n^1 \, du;$$

and on V_n^* we have $X_n^1 > 0$.

Remembering that on each face of the polyhedron S_n: $x^i = x^i(u)$ the Jacobians $X_n^i(u)$ are constant, we can express the above result in the following form. To each triangle $\tau_{j,k}$ of V_n we make correspond a vector $J_{j,k,n} = (J_{j,k,n}^1, J_{j,k,n}^2, J_{j,k,n}^3)$ defined by the equations $J_{j,k,n}^i \equiv \int_{\tau_{j,k}} X_n^i \, du$; this vector has the direction of the positive normal to the corresponding portion of the polyhedron S_n, and length equal to the area of that portion. Let the vector J_0 be defined by $J_0^i \equiv \int_q X_0^i \, du$.

Then the vector sum of the projections of $J_{j,k,n}$ on the vector (a_1, a_2, a_3) is $\sum_{j,k} a_\alpha J_{j,k,n}^\alpha$. This tends to $a_\alpha J_0^\alpha$; and moreover we can assume that for every j, k, and n the projection $a_\alpha J_{j,k,n}^\alpha$ has the same sign as $a_\alpha J_0^\alpha$.

5. **Fundamental lemma.** The following lemma is important in that Theorems I to IV are based on it:

LEMMA 2. *Let the function $f(x, X)$ be defined and satisfy conditions* (2.1) *on a point set A of x-space. Let S_0: $x^i = x_0^i(u)$, u on B_0, be a typical representation of a rectifiable surface interior to A. Let $f(x, X)$ satisfy the conditions*

(5.1a) *$f(x, X) \geqq 0$ for every x in some neighborhood of the points of S_0, and for every X;*

(5.1b) *$E(x_0(u), X_0(u), \overline{X}) \geqq 0$ for almost all points u in B_0 such that $X_0(u)$ is defined and $\| X_0(u) \| \neq 0$, and for all $\overline{X}$.*

Then for every sequence $\{S_n\}$ of rectifiable surfaces typically represented in the form S_n: $x^i = x_n^i(u)$, u on B_n, for which

$$\lim S_n = S_0, \quad \int_{B_n} \|X_n\| \, du \leqq N$$

(N a constant independent of n) it is true that

$$\liminf \int_{B_n} f(x_n, X_n) \, du \geqq \int_{B_0} f(x_0, X_0) \, du. \tag{5.2}$$

By § 3, it is sufficient to show that (5.2) holds for systems

$$S_0: \; x^i = x_0^i(u), \quad u \text{ on } Q, \tag{5.3}$$

$$S_n: \; x^i = x_n^i(u), \quad u \text{ on } Q, \tag{5.4}$$

in which the $x_n^i(u)$ are quasi-linear and

$$\lim_{n \to \infty} x_n^i(u) = x_0^i(u) \text{ uniformly on } Q. \tag{5.5}$$

By Remark 1, § 3, inequalities (5.1) continue to hold for the simplified system. We can also assume A to be closed and bounded; for if it is not, we restrict ourselves to a closed neighborhood of S_0, and for large enough values of n each S_n will lie in that neighborhood.

On the bounded closed set

$$[x \text{ in } A; \; X \text{ such that } \|X\| = 1] \tag{5.6}$$

the function $f(x, X)$ is continuous, hence uniformly continuous; hence for every $\varepsilon > 0$ there exists a $\delta > 0$ such that if $\|x_1, x_2\| \leqq 2\delta$,

$$|f(x_1, X) - f(x_2, X)| \leqq \frac{\varepsilon}{4N}. \tag{5.7}$$

Moreover the functions (5.3) are uniformly continuous on Q; hence there exists a $\gamma > 0$ such that if $\|u_1, u_2\| \leqq 2\gamma$,

$$\|x_0(u_1), x_0(u_2)\| \leqq \delta; \tag{5.8}$$

and also for all sufficiently large values of n

$$\|x_0(u), x_n(u)\| \leqq \delta \tag{5.9}$$

for every point u in Q.

The functions $f_{X^i}(x, X)$ are continuous on the set (5.6), hence are less in absolute value than some constant M. Since they are positively homogeneous of degree 0 in X, we have for every x in A and every X such that $\|X\| \neq 0$ the inequalities

$$|f_{X^i}(x, X)| \leqq M. \tag{5.10}$$

The functions $X_0^i(u)$ are measurable and bounded on Q; hence the set E_1 of points u of Q at which all three are approximately continuous has plane measure 1. On E_1 $f(x_0(u), X_0(u))$ is also approximately continuous. Letting u_0 be a point of E_1, we have for every sufficiently small square q with center at u_0 the inequalities

$$\left|\int_q X_0^i(u)\,du - \int_q X_0^i(u_0)\,du\right| \leqq \frac{\varepsilon\, m(q)}{24\,M}, \tag{5.11}$$

$$\left|\int_q f(x_0(u),\ X_0(u))\,du - \int_q f(x_0(u_0),\ X_0(u_0))\,du\right| \leqq \frac{\varepsilon\, m(q)}{4}. \tag{5.12}$$

Let β be a positive number so small that for every subset E of Q with measure $\leqq \beta$ we have

$$\int_E f(x_0(u),\ X_0(u))\,du \leqq \frac{\varepsilon}{4}. \tag{5.13}$$

By Vitali's theorem we can find a finite set of non-overlapping squares $q_1, \cdots, q_m$ lying in Q, having total measure $\sum_j m(q_j) \geqq 1-\beta$, each of side $\leqq \gamma$, and on each of which inequalities (5.11) and (5.12) are satisfied. Letting u_j be the center of square q_j we have by (5.12) and (5.13)

$$\left|\int_Q f(x_0(u),\ X_0(u))\,du - \sum_{j=1}^{m}\int_{q_j} f(x_0(u_j),\ X_0(u_j))\,du\right| \leqq \frac{\varepsilon}{2}. \tag{5.14}$$

On each square q_j we have by (5.8), (5.9), (5.7) and (2.1)

$$|f(x_n(u),\ X_n(u)) - f(x_0(u_j),\ X_n(u))| \leqq \|X_n(u)\| \cdot \frac{\varepsilon}{4N}; \tag{5.15}$$

and since $\sum_j \int_q \|X_n(u)\|\,du \leqq N$, this implies

$$\begin{aligned}\int_Q f(x_n,\ X_n)\,du &\geqq \sum_{j=1}^{m}\int_{q_j} f(x_n(u),\ X_n(u))\,du \\ &\geqq \sum_{j=1}^{m}\int_{q_j} f(x_0(u_j),\ X_n(u))\,du - \frac{\varepsilon}{4}.\end{aligned} \tag{5.16}$$

By Lemma 1 we can in each q_j define a sequence of subsets V_n^j such that

$$\lim_{n\to\infty}\int_{V_n^j} X_n^\alpha(u) f_{X^\alpha}(x_0(u_j),\ X_0(u_j))\,du = \int_{q_j} X_0^\alpha(u) f_{X^\alpha}(x_0(u_j),\ X_0(u_j))\,du. \tag{5.17}$$

If we define $\bar{X}_n$ by the equations

$$\begin{aligned}\bar{X}_n^i(u) &= X_n^i(u) \qquad (u \text{ on } V_n^j) \\ \bar{X}_n^i(u) &= 0 \qquad\quad (u \text{ on } q_j - V_n^j),\end{aligned} \tag{5.18}$$

we have from (5.17) and (5.1 a)

$$\int_{q_j} f(x_0(u_j),\ X_n(u))\ du \geqq \int_{q_j} f(x_0(u_j),\ \bar{X}_n(u))\ du, \tag{5.19}$$

$$\lim_{n\to\infty} \int_{q_j} (\bar{X}_n^\alpha(u) - X_0^\alpha(u))\, f_{X^\alpha}(x_0(u_j),\ X_0(u_j))\ du = 0. \tag{5.20}$$

For each square q_j we must have either $\| X_0(u_j)\| > 0$ or $\| X_0(u_j)\| = 0$. We consider these cases separately.

Case 1: $\| X_0(u_j)\| > 0$. In this case we must have $E(x_0(u_j),\ X_0(u_j),\ X) \geqq 0$ for all sets of numbers X; for if a set X exists such that at u_j

$$E(x_0(u_j),\ X_0(u_j),\ X) < 0, \tag{5.21}$$

then since $E(x_0(u),\ X_0(u),\ X)$ is approximately continuous at u_j, inequality (5.21) will continue to hold on a set of u's of positive measure, contrary to hypothesis. Let us write the identity

$$\begin{aligned}\int_{q_j} f(x_0(u_j),\ \bar{X}_n(u))\ du = &\int_{q_j} E(x_0(u_j),\ X_0(u_j),\ \bar{X}_n(u))\ du \\ &+ \int_{q_j} (\bar{X}_n^\alpha(u) - X_0^\alpha(u))\ f_{X^\alpha}(x_0(u_j),\ X_0(u_j))\ du \\ &+ \int_{q_j} (X_0^\alpha(u) - X_0^\alpha(u_j))\, f_{X^\alpha}(x_0(u_j),\ X_0(u_j))\ du \\ &+ \int_{q_j} f(x_0(u_j),\ X_0(u_j))\ du.\end{aligned} \tag{5.22}$$

As we have just seen, the first integral on the right is $\geqq 0$. For all sufficiently large values of n the second integral has absolute value at most $\varepsilon\, m(q_j)/8$, by (5.20); and by (5.11) and (5.10) the absolute value of the third integral is at most $\varepsilon\, m(q_j)/8$. This, together with (5.19), gives us

$$\int_{q_j} f(x_0(u_j),\ X_n(u))\ du \geqq \int_{q_j} f(x_0(u_j),\ X_0(u_j))\ du - \frac{\varepsilon\, m(q_j)}{4}. \tag{5.23}$$

We turn now to Case 2: $\| X_0(u_j)\| = 0$. For this case inequality (5.23) is immediate, since $f(x_0(u_j),\ X_0(u_j)) = 0$, and by hypothesis the left member of (5.23) is $\geqq 0$.

Now from (5.16), (5.23), and (5.14) we conclude that for all sufficiently large values of n we have

$$\begin{aligned}\int_Q f(x_n,\ X_n)\ du &\geqq \sum_j \int_{q_j} f(x_0(u_j),\ X_n(u))\ du - \frac{\varepsilon}{4} \\ &\geqq \sum_j \int_{q_j} f(x_0(u_j),\ X_0(u_j))\ du - \frac{\varepsilon}{2} \\ &\geqq \int_Q f(x_0,\ X_0)\ du - \varepsilon;\end{aligned}$$

hence $\liminf \int_Q f(x_n, X_n)\, du \geqq \int_Q f(x_0, X_0)\, du$, and inequality (5.2) is valid.

6. **Invariance. Semi-continuity on restricted classes of surfaces.** In order to justify the use of the notation $F(S)$ we must prove

THEOREM I. *Let $f(x, X)$ satisfy conditions (2.1) on a point set A of x-space. Let the rectifiable surface S be interior to A, and have the two typical representations*

$$S: \quad x^i = x^i(u), \quad u \text{ on } B, \tag{6.1}$$

$$S: \quad x^i = \bar{x}^i(\bar{u}), \quad \bar{u} \text{ on } \bar{B}. \tag{6.2}$$

Then

$$\int_{\bar{B}} f(\bar{x}, \bar{X})\, d\bar{u} = \int_B f(x, X)\, du. \tag{6.3}$$

First, let us suppose that conditions (5.1) are satisfied on S. We define, for every positive integer n, $B_n = \bar{B}$, and set $x_n^i(u) = \bar{x}^i(u)$, u on $\bar{B}$, so that $\|S, S_n\| = 0$ for every n. Then by Lemma 2 we have

$$\int_{\bar{B}} f(\bar{x}, \bar{X})\, d\bar{u} \geqq \int_B f(x, X)\, du.$$

Interchanging the roles of (6.1) and (6.2), we obtain (6.3).

Suppose now that conditions (5.1) are not fulfilled on S. Define $\varphi(x, X) \equiv \|X\|$; we easily calculate that the corresponding E-function

$$E_\varphi(x, X, \bar{X}) \equiv \|\bar{X}\| \cdot (1 - \cos\alpha),$$

where α is the angle between the directions X and $\bar{X}$. If we denote by A^* a bounded closed subset of A to which S is interior, we find[15] that on A^*

$$|E(x, X, \bar{X})| \leqq M \cdot \|\bar{X}\| \cdot (1 - \cos\alpha),$$

where M is a constant and α the angle between X and $\bar{X}$. And by (2.1), if N be the upper bound of $|f(x, \bar{X})|$ for all x on A^* and all $\bar{X}$ such that $\|\bar{X}\| = 1$ we have $|f(x, X)| \leqq N \cdot \|X\|$.

Choosing H as a constant greater than M and N, we find that the function $f(x, X) + H\varphi(x, X)$ satisfies conditions (5.1) on S. Consequently

$$\int_{\bar{B}} f(\bar{x}, \bar{X})\, d\bar{u} + H\int_{\bar{B}} \varphi(\bar{x}, \bar{X})\, d\bar{u} = \int_{\bar{B}} f(x, X)\, du + H\int_B \varphi(x, X)\, du.$$

[15] G. A. Bliss, *The Weierstrass E-Function for Problems of the Calculus of Variations in Space*, Trans. Am. Math. Soc., vol. 15 (1914), pp. 369–378.

In this paper Bliss considers curves, not surfaces, in three-space; but by a simple change of notation we find that his results apply equally well to our case.

But since $\varphi(x, X)$ itself satisfies conditions (5.1) on S, the second term on the right is equal to the second term on the left; hence equation (6.3) is valid, even if conditions (5.1) are not satisfied.

Let us denote the area, in the sense of Lebesgue, of the surface S by the symbol $L(S)$. It is now easy to prove

THEOREM II. *If the rectifiable surface S be typically represented in the form S: $x^i = x^i(u)$, u on B, then*

$$L(S) = \int_B \| X \| \, du. \tag{6.4}$$

The integrand $f(x, X) \equiv \| X \|$ satisfies conditions (5.1) on every surface. By a direct calculation we find that equation (6.4) holds for typically represented polyhedra; and the area is a lower semi-continuous functional, as follows immediately from its definition.

Therefore if $\{\Pi_n\}$ be a sequence of polyhedra typically represented in the form Π_n: $x^i = x^i_n(u)$, u on B_n, of uniformly bounded area, and approaching S as limit, we can write

$$L(\Pi_n) = \int_{B_n} \| X_n \| \, du, \qquad (n = 1, 2, \cdots), \tag{6.5}$$

$$\liminf \int_{B_n} \| X_n \| \, du \geqq \int_B \| X \| \, du, \tag{6.6}$$

$$\liminf L(\Pi_n) \geqq L(S). \tag{6.7}$$

By the definition of area we can find such a sequence of polyhedra for which the equality sign holds in (6.7), from which it follows that $\int_B \| X \| \, du \leqq L(S)$. On the other hand we can by Remark 2 of § 3 find a sequence of polyhedra for which the equality holds in (6.6), from which it follows that $\int_B \| X \| \, du \geqq L(S)$. These two inequalities together establish equation (6.4).

Theorems I and II, together with Lemma 2, establish

THEOREM III. *Let the function $f(x, X)$ be defined and satisfy conditions* (2.1) *on a point set A of x-space. Let S_0: $x^i = x^i_0(u)$, u on B_0, be a typical representation of a rectifiable surface interior to A. Let $f(x, X)$ satisfy the conditions*

a) *$f(x, X) \geqq 0$ for every x in some neighborhood of the points of S_0, and for every X;*

b) *$E(x_0(u), X_0(u), \bar{X}) \geqq 0$ for almost all points u in B such that $X_0(u)$ is defined and $\| X_0(u) \| \neq 0$, and for all $\bar{X}$.*

Then for every positive number N the functional $F(S) \equiv \int f(x, X) \, du$ is lower semi-continuous at S_0 with respect to the class of all rectifiable surfaces of area $\leqq N$.

As an immediate consequence of Theorem III we have

THEOREM IV. *If $F(S)$ be positive semi-definite, positive quasi-regular on A, and $\mathfrak{R}$ be a class of rectifiable surfaces $[S]$ interior to A and of uniformly bounded area, then $F(S)$ is lower semi-continuous on $\mathfrak{R}$.*

7. **Semi-continuity with respect to all rectifiable surfaces.** In this sections we shall develop conditions which are sufficient to insure lower semi-continuity on the class of all rectifiable surfaces interior to A, without restriction as to area. We find it convenient to introduce a new notation; a subscript u at the left of a vector shall indicate that it is a unit vector; that is, if we write ${}_uX$ we shall understand that we are considering a vector such that $\|{}_uX\| = 1$. Moreover, let $\bar{X}$ be any vector such that $\|\bar{X}\| \neq 0$; by the symbol ${}_u\bar{X}$ we shall designate the unit vector $\bar{X}/\|\bar{X}\|$. We have then $\bar{X}^i = ({}_u\bar{X}^i)(\|\bar{X}\|)$.

A first and simple theorem is

THEOREM V. *If in the hypotheses of Theorem III we replace* (5.1a) *by the condition*

(7.1) $f(x_0, {}_uX) > 0$ *for every x_0 on S_0 and every ${}_uX$, then $F(S)$ is lower semi-continuous at S_0 with respect to the class of all rectifiable surfaces interior to A.*

By the Heine-Borel theorem we can find a neighborhood W of S_0 such that for every x in W and every ${}_uX$ we have $f(x, {}_uX) \geqq m > 0$. Consider a sequence $\{S_n\}$ of rectifiable surfaces S_n: $x^i = x_n^i(u)$, u on B_n, tending to S_0. For sufficiently large values of n the surface S_n will lie in W, so that by (2.1)

$$(7.2)\qquad \int_{B_n} f(x_n, X_n)\, du = \int_{B_n} f(x_n, {}_uX_n)\, \|X_n\|\, du \geqq m \cdot \int_{B_n} \|X_n\|\, du.$$

If it were now true that $\liminf F(S_n) < F(S_0)$, we would be able to select a subsequence of $\{S_n\}$ (for which we retain the same notation) such that

$$(7.3)\qquad \lim F(S_n) < F(S_0).$$

By (7.2) this implies that $\int_{B_n} \|X_n\|\, du$ is uniformly bounded, so that by Theorem III the inequality (7.3) is impossible.

As a corollary we have

THEOREM VI. *If $F(S)$ is positive definite, positive quasi-regular on A, it is lower semi-continuous on the class of all rectifiable surfaces interior to A.*

We now proceed to prove

THEOREM VII. *Let the function $f(x, X)$ be defined and satisfy conditions* (2.1) *on a point set A of x-space. Let S_0 be a rectifiable surface typically*

represented in the form $S_0 : x^i = x_0^i(u)$, u *on* B_0, *interior to* A. *Let* $f(x, X)$ *satisfy the conditions:*

(7.4) $f(x, X) \geqq 0$ *for every* X *and for every* x *in a neighborhood of the point-set constituting* S_0;

(7.5) $E(x_0(u), X_0(u), {}_u\bar{X}) > 0$ *for almost every point* u *of* B_0 *at which* $X_0(u)$ *is defined and* $\| X_0(u) \| > 0$, *and for every* ${}_u\bar{X} \neq {}_uX_0(u)$.

Then $F(S)$ *is lower semi-continuous at* S_0 *with respect to the class of all rectifiable surfaces interior to* A.

As before, we need only to prove $\liminf F(S_n) \geqq F(S_0)$, with the assumptions that B_0 is the unit square Q, and the surfaces S_n are polyhedra typically represented in the form S_n: $x^i = x_n^i(u)$, u on Q, with $\lim x_n^i(u) = x_0^i(u)$ uniformly on Q. By Remark 1, § 3, we see that hypotheses (7.4) and (7.5) continue to hold after the simplifications.

Let ε be an arbitrary positive number less than 1. Defining E_1 as the set of all points u of Q at which $X_0(u)$ is defined, (7.5) holds, and $f(x_0(u), X_0(u)) \geqq \varepsilon/8$, we have

$$(7.6)\quad F(S_0) = \int_{E_1} f(x_0, X_0)\, du + \int_{Q-E_1} f(x_0, X_0)\, du \leqq \int_{E_1} f(x_0, X_0)\, du + \frac{\varepsilon}{8}.$$

Denote by $\alpha(u)$ the angle between the vectors $X_0(u)$ and $f_{X^0}(x_0(u), X_0(u))$; on E_1 we have

$$\frac{\varepsilon}{8} \leqq f(x_0, X_0) = X_0^\alpha f_{X^\alpha}(x_0, X_0) = \| X_0 \| \cdot \| f_X(x_0, X_0) \| \cdot \cos \alpha(u),$$

in which $\| f_X(x_0, X_0) \|$ denotes $[f_{X^\alpha}(x_0, X_0) f_{X^\alpha}(x_0, X_0)]^{1/2}$. Since the coefficient of $\cos \alpha(u)$ is bounded on Q, there exists a constant $\vartheta > 0$ such that on E_1

$$(7.7)\qquad 0 \leqq \alpha(u) \leqq \frac{\pi}{2} - 2\vartheta.$$

Moreover, $\| f_X(x_0, X_0) \| > L$, where L is a positive constant.

Since $f(x_0(u), X_0(u))$ is summable on Q, there exists a $\beta > 0$ such that for every subset E of Q having measure $m(E) \leqq \beta$ it is true that

$$(7.8)\qquad \int_E f(x_0(u), X_0(u))\, du \leqq \frac{\varepsilon}{8}.$$

Let x be any point in A, and ${}_uX$ any unit vector. As the variable vector ${}_u\bar{X}$ ranges over all unit vectors such that the angle between ${}_uX$ and ${}_u\bar{X}$ is at least ϑ, the function $E(x, {}_uX, {}_u\bar{X})$ has a certain lower bound which we denote by $\mu(x, {}_uX)$. We easily see that $\mu(x, {}_uX)$ is a continuous function of its arguments; hence on E_1 the function $\mu(x_0(u), {}_uX_0(u))$ is measurable. Moreover, it is by (7.5) positive on E_1, hence for a certain $m > 0$ the subset E_2 of E_1 on which $\mu(x_0, {}_uX_0) \geqq m$ has measure at least equal to $m(E_1) - \beta$.

From (7.6), (7.8) we conclude that

$$\int_{E_2} f(x_0, X_0)\,du \geqq \int_{E_1} f(x_0, X_0)\,du - \frac{\varepsilon}{8} \geqq F(S_0) - \frac{\varepsilon}{4}. \tag{7.9}$$

Almost every point of E_2 is a point of approximate continuity of X_0^1, X_0^2, and X_0^3; and at every such point u_0 we have for all sufficiently small squares q with center at u_0

$$\left| \int_q f(x_0(u), X_0(u))\,du - \int_q f(x_0(u_0), X_0(u_0))\,du \right| \leqq \frac{\varepsilon\, m(q)}{8}, \tag{7.10}$$

$$\left| \int_q X_0^i(u)\,du - \int_q X_0^i(u_0)\,du \right| \leqq \frac{\varepsilon\, m(q)}{24\,M}, \tag{7.11}$$

where M is defined by inequalities (5.10).

On the bounded closed set [x in A; ${}_uX$ such that $\|{}_uX\| = 1$] the function $f(x, {}_uX)$ is uniformly continuous; hence there exists a $\delta > 0$ such that if $\|x_1, x_2\| \leqq 2\delta$,

$$|f(x_1, {}_uX) - f(x_2, {}_uX)| \leqq \text{smaller } \left\{ \frac{m}{2},\ \frac{\varepsilon L \sin\vartheta}{16(H+1)} \right\}, \tag{7.12}$$

where H is an upper bound for $f(x_0(u), X_0(u))$ on Q. Also, by the uniform continuity of $x_0(u)$, there exists a $\gamma > 0$ such that if $\|u_1, u_2\| \leqq \gamma$, then

$$\|x_0(u_1), x_0(u_2)\| \leqq \delta; \tag{7.13}$$

and since $\lim x_n^i(u) = x_0^i(u)$ uniformly on Q, we have for all sufficiently large values of n

$$\|x_n(u), x_0(u)\| \leqq \delta \text{ for all } u \text{ on } Q. \tag{7.14}$$

By Vitali's theorem we can find a finite set of squares $q_1, \cdots, q_h$, each of side $\leqq \gamma$, whose centers u_j are points of E_2 at which the X_0^i are approximately continuous, such that $m(E_2 - \sum q_j) \leqq \beta$, and such that (7.10) and (7.11) are satisfied. Then by (7.10), (7.4), (7.8) and (7.9) we have

$$\begin{aligned} &\sum_{j=1}^{h} \int_{q_j} f(x_0(u_j), X_0(u_j))\,du \\ &\quad \geqq \sum_{j=1}^{h} \int_{q_j} f(x_0(u), X_0(u))\,du - \frac{\varepsilon}{8} \\ &\quad \geqq \int_{E_2} f(x_0(u), X_0(u))\,du - \frac{\varepsilon}{4} \\ &\quad \geqq F(S_0) - \frac{\varepsilon}{2}. \end{aligned} \tag{7.15}$$

By Lemma 1 we can find a sequence of subsets V_n^j of q_j such that

$$\lim_{n\to\infty} \int_{V_n^j} X_n^\alpha(u) f_{X^\alpha}(x_0(u_j), X_0(u_j))\, du = \int_{q_j} X_0^\alpha(u) f_{X^\alpha}(x_0(u_j), X_0(u_j))\, du; \tag{7.16}$$

and by the remark after Lemma 1, we can assume that the integrand on the left is always of the same sign as the integral on the right. We define

$$\begin{aligned} \bar{X}_n^i(u) &\equiv X_n^i(u), \quad u \text{ on } V_n^j, \\ \bar{X}_n^i(u) &\equiv 0, \quad u \text{ on } q_j - V_n^j. \end{aligned} \tag{7.17}$$

Then by (7.4) we see that

$$F(S_n) \geqq \sum_{j=1}^{h} \int_{q_j} f(x_n(u), \bar{X}_n(u))\, du. \tag{7.18}$$

Let $T_{1,j,n}$ be the subset of V_n^j on which the angle between $X_0(u_j)$ and $X_n(u)$ is $\leqq \vartheta$, and $T_{2,j,n}$ the remainder of the square q_j. On $T_{1,j,n}$ the angle between $X_0(u_j)$ and the vector $f_{X^i}(x_0(u_j), X_0(u_j))$ is by (7.7) less than $\pi/2 - 2\vartheta$, and the angle between $X_n(u)$ and $X_0(u_j)$ is $\leqq \vartheta$, so that the angle between f_{X^i} and $X_n(u)$ is $\leqq \pi/2 - \vartheta$; whence

$$\begin{aligned} X_n^\alpha(u) f_{X^\alpha}(x_0(u_j), X_0(u_j)) &\geqq \|X_n(u)\| \cdot \|f_X\| \cdot \cos\left(\frac{\pi}{2} - \vartheta\right) \\ &= \|X_n(u)\| \cdot \|f_X\| \cdot \sin\vartheta. \end{aligned} \tag{7.19}$$

This with (7.16) implies that for all sufficiently large values of n

$$\begin{aligned} &2\int_{q_j} X_0^\alpha(u) f_{X^\alpha}(x_0(u_j), X_0(u_j))\, du \\ &\geqq \int_{V_n^j} X_n^\alpha(u) f_{X^\alpha}(x_0(u_j), X_0(u_j))\, du \\ &\geqq \int_{T_{1,j,n}} X_n^\alpha(u) f_{X^\alpha}(x_0(u_j), X_0(u_j))\, du \\ &\geqq \sin\vartheta \cdot L \cdot \int_{T_{1,j,n}} \|X_n\|\, du, \end{aligned} \tag{7.20}$$

or using (7.10) and (7.11),

$$\begin{aligned} \int_{T_{1,j,n}} \|X_n\|\, du &\leqq \left(\frac{2}{L\sin\vartheta}\right) \int_{q_j} X_0^\alpha(u) f_{X^\alpha}(x_0(u_j), X_0(u_j))\, du \\ &= \left(\frac{2}{L\sin\vartheta}\right) \left\{ \int_{q_j} [X_0^\alpha(u) - X_0^\alpha(u_j)] f_{X^\alpha}(x_0(u_j), X_0(u_j))\, du \right. \\ &\qquad \left. + \int_{q_j} X_0^\alpha(u_j) f_{X^\alpha}(x_0(u_j), X_0(u_j))\, du \right\} \\ &\leqq \left(\frac{2}{L\sin\vartheta}\right) \left\{ \frac{\varepsilon\, m(q_j)}{8} + \int_{q_j} f(x_0(u_j), X_0(u_j))\, du \right\} \\ &\leqq \frac{2(H+1)\, m(q_j)}{L\sin\vartheta}, \end{aligned} \tag{7.21}$$

recalling in the last step that H is by definition an upper bound for $f(x_0(u), X_0(u))$ on Q.

On $T_{2,j,n}$ we have by the definition of E_2

$$E(x_0(u_j), X_0(u_j), X_n(u)) \geqq m \cdot \| X_n(u) \|. \tag{7.22}$$

Let us now consider the identity

$$\begin{aligned}
&\int_{q_j} f(x_n(u), \bar{X}_n(u))\, du - \int_{q_j} f(x_0(u_j), X_0(u_j))\, du \\
&= \int_{T_{1,j,n}} + \int_{T_{2,j,n}} [f(x_n(u), {}_u\bar{X}_n(u)) - f(x_0(u_j), {}_u\bar{X}_n(u))] \cdot \| \bar{X}_n(u) \|\, du \\
&\quad + \int_{T_{1,j,n}} + \int_{T_{2,j,n}} E(x_0(u_j), X_0(u_j), {}_u\bar{X}_n(u)) \| \bar{X}_n(u) \|\, du \\
&\quad + \int_{q_j} [\bar{X}_n^\alpha(u) - X_0^\alpha(u)] f_{X^\alpha}(x_0(u_j), X_0(u_j))\, du \\
&\quad + \int_{q_j} [X_0^\alpha(u) - X_0^\alpha(u_j)] f_{X^\alpha}(x_0(u_j), X_0(u_j))\, du,
\end{aligned} \tag{7.23}$$

wherein we agree to set ${}_u\bar{X}_n(u) = (1, 0, 0)$ when $\| \bar{X}_n(u) \| = 0$. Denote the six integrals on the right by $i_1, i_2, \cdots, i_6$. Then by (7.12), (7.13), (7.14), (7.21)

$$i_1 \geqq -\frac{\varepsilon L \sin \vartheta}{16(H+1)} \int_{T_{1,j,n}} \| \bar{X}_n(u) \|\, du \geqq -\frac{\varepsilon\, m(q_j)}{8}.$$

By (7.12), (7.13), (7.14),

$$i_2 \geqq -\frac{m}{2} \int_{T_{2,j,n}} \| \bar{X}_n(u) \|\, du,$$

and by (7.22)

$$i_4 \geqq m \int_{T_{2,j,n}} \| \bar{X}_n(u) \|\, du.$$

By (7.5), $i_3 \geqq 0$. For all sufficiently large values of n we have by (7.16) and (7.17) $i_5 \geqq -\varepsilon\, m(q_j)/4$, and by (7.10) (with (5.10)) we have $i_6 \geqq -\varepsilon\, m(q_j)/8$.

Summing over all squares q_j, we find that the left member of (7.23) is $\geqq -\varepsilon/2$. This, with (7.18) and (7.15), proves that for all sufficiently large values of n $F(S_n) \geqq F(S_0) - \varepsilon$; so that $\liminf F(S_n) \geqq F(S_0)$, and our theorem is established.

An immediate corollary of Theorem VII is

THEOREM VIII. *If for every x in A, every unit vector ${}_uX$, and every unit vector ${}_uX^* \neq {}_uX$ we have*

$$f(x, {}_uX) \geqq 0, \quad E(x, {}_uX, {}_uX^*) > 0,$$

then $F(S)$ is lower semi-continuous on the class of all rectifiable surfaces interior to A.

8. **A comparison theorem.** Suppose that $f(x, X)$ is defined and satisfies conditions (2.1) on a point set A. For brevity we denote $f_{X^iX^k}$ by f_{ik}; we find as a consequence of the homogeneity relations

$$X^\alpha f_{\alpha k}(x, X) \equiv 0, \tag{8.1}$$

so that the quadratic form $X^\alpha X^\beta f_{\alpha\beta}(x, X)$ vanishes identically. Also, $f_{ik}(x, X)$ is positively homogeneous of degree -1 in X. As an immediate consequence of a formula of Bliss[16] we find that the functional $F(S)$ is positive quasi-regular if and only if $\bar{X}^\alpha \bar{X}^\beta f_{\alpha\beta}(x, X) \geqq 0$ for all x in A, all $\bar{X}$, and all X such that $\|X\| \neq 0$. We will say that $F(S)$ is *positive regular* on A if for every pair of orthogonal unit vectors ${}_uX$, ${}_u\bar{X}$ (that is, every pair such that ${}_uX^\alpha\, {}_u\bar{X}^\alpha = 0$) and every x in A we have

$${}_u\bar{X}^\alpha\, {}_u\bar{X}^\beta f_{\alpha\beta}(x, {}_uX) > 0. \tag{8.2}$$

From this it follows that if $F(S)$ is positive regular on A, then for every pair of vectors X, $\bar{X}$ for which $\|X\| \neq 0$, $\|\bar{X}\| \neq 0$, and ${}_u\bar{X} \neq {}_uX$, we have $\bar{X}^\alpha \bar{X}^\beta f_{\alpha\beta}(x, X) > 0$. For we can find a unit vector ${}_uX_0$ orthogonal to X such that ${}_u\bar{X}^i = a \cdot {}_uX_0^i + b \cdot X^i$, $a \neq 0$; then by (8.1)

$$\begin{aligned} \bar{X}^\alpha \bar{X}^\beta f_{\alpha\beta}(x, X) &= (\|\bar{X}\|^2/\|X\|)\, {}_u\bar{X}^\alpha\, {}_u\bar{X}^\beta f_{\alpha\beta}(x, {}_uX) \\ &= (\|\bar{X}\|^2/\|X\|)\, [a^2\, {}_uX_0^\alpha\, {}_uX_0^\beta f_{\alpha\beta}(x, {}_uX)] > 0. \end{aligned}$$

Consequently if $F(S)$ be positive regular on A, it is *a fortiori* positive quasi-regular on A.

Using the notation $F(S) \equiv \int f(x, X)\, du$, $G(S) \equiv \int g(x, X)\, du$, we now prove

THEOREM IX. *Let $f(x, X)$ and $g(x, X)$ be defined and satisfy conditions (2.1) on a point set A of x-space. Let the rectifiable surface S_0 be interior to A, and on the point set constituting S_0 let $f(x, X)$ be positive definite and positive regular. Then for every sequence $\{S_n\}$ of rectifiable surfaces such that*

$$\lim S_n = S_0, \tag{8.3}$$

$$\lim F(S_n) = F(S_0) \tag{8.4}$$

it is true that

$$\lim G(S_n) = G(S_0). \tag{8.5}$$

[16] loc. cit. [15].

For each point x of A we define $\mu(x)$ as the greatest lower bound, for all pairs of orthogonal unit vectors X, $\bar{X}$, of the expression $\bar{X}^\alpha \bar{X}^\beta f_{\alpha\beta}(x, X)$. We readily see that $\mu(x)$ is continuous. On the bounded closed set of points constituting S_0 we have by (8.2) $\mu(x) > 0$, hence $\mu(x) \geqq 2m > 0$. Hence we can find a closed neighborhood W of S_0 on which $\mu(x) \geqq m$; and for all sufficiently large values of n S_n must lie in W.

Utilizing the formula of Bliss we have for every x on W, every $\bar{X}$, and every X such that $\|X\| \neq 0$, the inequality

$$E_f(x, X, \bar{X}) \geqq m \cdot \|\bar{X}\| \cdot (1 - \cos\alpha),$$

where α is the angle between X and $\bar{X}$, and E_f is the E-function formed for $f(x, X)$. By the same formula we have

$$|E_g(x, X, \bar{X})| \leqq M \cdot \|\bar{X}\| \cdot (1 - \cos\alpha),$$

where M is some constant. Hence for every constant $H \geqq M/m$ both of the functionals $H \cdot F(S) + G(S)$ and $H \cdot F(S) - G(S)$ will be positive quasi-regular on W. By a similar and simpler argument we see that if H be large enough both of these functionals will also be positive definite on S_0. Hence by Theorem VII we will have

$$\begin{aligned} \liminf\,[H \cdot F(S_n) + G(S_n)] &\geqq H \cdot F(S_0) + G(S_0), \\ \liminf\,[H \cdot F(S_n) - G(S_n)] &\geqq H \cdot F(S_0) - G(S_0). \end{aligned}$$

By virtue of equation (8.4) these inequalities imply

$$\begin{aligned} \liminf \quad G(S_n) &\geqq \quad G(S_0), \\ \liminf\,[-G(S_n)] &\geqq -G(S_0); \end{aligned}$$

from which (8.5) follows immediately.

COROLLARY. *If $f(x, X)$ be defined and satisfy conditions* (2.1) *on a point set A of x-space, and S_0 be a rectifiable surface interior to A, then for every sequence $\{S_n\}$ of rectifiable surfaces such that*

$$\lim S_n = S_0, \quad \lim L(S_n) = L(S_0),$$

it is true that

$$\lim F(S_n) = F(S_0).$$

For by Theorem IV we have $L(S) = \int \|X\| \cdot du$, and we readily verify that this functional is on the whole of x-space both positive definite and positive regular.

9. **Continuous functionals.** If we compare Theorems III, IV, VII and VIII with their analogues[17] for curves, we find as the most important

[17] Tonelli, *Fondamenti di Calcolo delle Variazione*, vol. I, §§ 108, 106, 96.

distinction the added hypothesis that $f(x, X) \geqq 0$. As a matter of fact, this added condition is actually necessary in order that $F(S)$ be lower semi-continuous on the class of all rectifiable surfaces interior to A. For suppose that for a certain (x, X), with x interior to A, we have $f(x, X) = -k < 0$; we can without loss of generality assume $x = (0, 0, 0)$, $X = (0, 0, 1)$. Then the sequence $\{S_n\}$ of surfaces S_n: $x^1 = n^{-1} u^1 \cos n^2 u^2$, $x^2 = n^{-1} u^1 \sin n^2 u^2$, $x^3 = 0$,

$$0 \leqq u^1 \leqq 1, \quad 0 \leqq u^2 \leqq 2\pi,$$

approach the limit surface S_0: $x^i = 0$. Each S_n is rectifiable, and we easily calculate $L(S_n) = \pi$ for every n. For large enough values of n S_n lies in A; and

$$\lim F(S_n) = -k\pi < 0 = F(S_0).$$

As a consequence of this we find that the only functional which is continuous on the class of all rectifiable surfaces, or even on the class of all rectifiable surfaces of area less than some constant N, is the trivial one $F(S) \equiv 0$; for to have both upper and lower semi-continuity we must have

$$f(x, X) \leqq 0, \quad f(x, X) \geqq 0.$$

However, if we limit ourselves to the class of rectifiable surfaces bounded by a given fixed curve of finite length the condition $f(x, X) \geqq 0$ is no longer necessary; for in fact for any three constants a_i the functional $F(S) = \int a_\alpha X^\alpha \, du$ is continuous on such a class of surfaces. For such a class of surfaces we have as a necessary condition for continuity $E \geqq 0$, $E \leqq 0$; hence $f(x, X) = a_\alpha(x) X^\alpha$, where the $a_i(x)$ are continuous functions of the x^i. It would be interesting to investigate more fully the conditions for continuity on such a class of surfaces.

HARVARD UNIVERSITY,
May 13, 1931.

SOME APPLICATIONS OF THE CALCULUS OF VARIATIONS TO RIEMANNIAN GEOMETRY.[1]

By I. J. Schoenberg.[2]

In 1855 O. Bonnet, using Jacobi's variational equations for geodesics on a surface and Sturm's comparison theorem for linear second order differential equations, established a theorem relating the curvature of an ovaloid surface to its size: *If at every point of an ovaloid surface V_2, its curvature K is $\geqq 1/A^2$ (A constant), then the distance between any two points on V_2 is $< \pi A$.*[3]

It is the purpose of this paper to generalize this theorem to Riemannian geometry. The Theorem 4.1 here proved is based on the assumption of the existence of an absolute shortest distance in V_n between any two points of V_n which is furnished by a geodesic arc joining those points. The application of recent existence theorems for the absolute minimum makes it possible however to derive from Theorem 4.1 more special and precise results like Bonnet's theorem.

The method of proof here set forth is similar to Bonnet's with the following changes: We do not work with Levi-Civita's variational equations which generalize Jacobi's equation for geodesics, but directly with the second variation of the length integral of V_n. This second variation is set up by a method first used by L. Berwald.[4] The use of Levi-Civita's geodesic coördinates[5] turns out to be very convenient (§ 2). Accordingly, Sturm's comparison theorem is replaced by a comparison theorem for the second variation. This theorem, a special case of a comparison theorem of Morse,[6] will be established quite independently with the help only of classical results of the calculus of variations (§ 1).

---

[1] Received July 1st, 1931. Presented to the American Mathematical Society June 13, 1931.

[2] Fellow of the International Education Board.

[3] O. Bonnet, *Sur quelques propriétés des lignes géodésiques*, Comptes Rendus 40 (1855), pp. 1311–1313. See also W. Blaschke, *Differentialgeometrie* I, Berlin 1930, § 100.

[4] L. Berwald, *Una forma invariante della seconda variazione*, Rendiconti dei Lincei, serie IV, vol. 7 (1928), pp. 301–306. Compare J. L. Synge, *The first and second variations of the length integral in Riemann space*, Proc. of the London Math. Society, 2nd series, vol. 25 (1926), pp. 247–264.

[5] T. Levi-Civita, *Sur l'écart géodésique*, Math. Annalen, Bd. 97 (1927), pp. 291–320. For more recent papers on this subject see E. Bortolotti, *Sulle coordinate geodetiche lungo una linea*, Rend. dei Lincei, vol. 10 (1929), pp. 486–492.

[6] M. Morse, *A generalization of the Sturm separation and comparison theorems in n-space*, Math. Annalen, Bd. 103, pp. 52–69.

The problem of the shortest distance in V_n is also discussed. A sufficiency theorem of Bliss for the general parametric problem of the calculus of variations will easily solve this problem (§ 5).

1. **A comparison theorem for the second variation.** We consider the quadratic form in u_i and v_i, $(i = 1, 2, \cdots, n)$,

$$(1.1) \qquad 2\,\Omega(x, u, v) = P_{ik}(x)u_i u_k + 2\,Q_{ik}(x)u_i v_k + R_{ik}(x)v_i v_k,$$

the coefficients $P_{ik}(x)$, $Q_{ik}(x)$, and $R_{ik}(x)$ being functions of x of class C'[7] for $x_1 \leqq x \leqq x_2$, with $P_{ik} = P_{ki}$, $R_{ik} = R_{ki}$ and

$$(1.2) \qquad R_{ik}(x)v_i v_k > 0 \quad \text{for} \quad x_1 \leqq x \leqq x_2, \quad v_i v_i > 0.$$

We set up the integral

$$(1.3) \quad I(\eta) = \int_{x_1}^{x_2} (P_{ik}\eta_i \eta_k + 2\,Q_{ik}\eta_i \eta_k' + R_{ik}\eta_i' \eta_k')\,dx = \int_{x_1}^{x_2} 2\,\Omega(x, \eta, \eta')\,dx.$$

The problem of minimizing $I(\eta)$ in the class of sets $\eta_i = \eta_i(x)$ $(i = 1, 2, \cdots, n)$, $\eta_i(x)$ being functions of class D'[8] on $x_1 x_2$, with $\eta_i(x_1) = \eta_i(x_2) = 0$ (we shall call such a set an admissible one), leads to the system of second order linear homogeneous differential equations

$$(1.4) \qquad \frac{d}{dx}\Omega\eta_i' - \Omega\eta_i = 0, \qquad (i = 1, 2, \cdots, n),$$

the Euler equations for $I(\eta)$. The set $\eta_i \equiv 0$, $x_1 \leqq x \leqq x_2$, is an admissible one, satisfies the system (1.4) and gives $I(\eta)$ the value zero.

The calculus of variations provides conditions under which $I(0) = 0$ is the minimum value of $I(\eta)$ in the class of admissible sets $\eta_i = \eta_i(x)$. The extremal arc $\eta_i \equiv 0$, $x_1 \leqq x \leqq x_2$, satisfies Legendre's condition in the strict sense, because of (1.2), which insures $|R_{ik}(x)| \neq 0$ on $x_1 x_2$. Jacobi's condition deals with conjugate points: The point ξ on $x_1 x_2$ is said to be conjugate to x_1, if there is a set $u_i = u_i(x)$ of class C' and not all $u_i \equiv 0$ on $x_1 \xi$, satisfying (1.4) and $u_i(x_1) = u_i(\xi) = 0$. Jacobi's condition is satisfied in the strict sense, in case there is no such ξ on the interval $x_1 x_2$.

In the calculus of variations the following lemma is proved:[9]

LEMMA 1.1. *If there is no point on $x_1 x_2$ conjugate to x_1, then one always has $I(\eta) > 0$, for every set $\eta_i = \eta_i(x)$ of class D' on $x_1 x_2$, with $\eta_i(x_1) = \eta_i(x_2) = 0$ and not all η_i identically zero on $x_1 x_2$.*

[7] A function is of class C' if it is continuous and has a continuous first derivative.

[8] A function is of class D' if it is continuous, and if the interval $x_1 x_2$ can be devided into a finite number of subintervals, such that in each subinterval its derivative should be continuous (O. Bolza).

[9] See J. Hadamard, *Calcul des Variations*, I, Paris 1910, § 288.

We shall prove now the following

COMPARISON THEOREM 1.1. *Besides $I(\eta)$ given by* (1.1), *condition* (1.2) *being retained, we consider the integral*

$$I^*(\eta) = \int_{x_1}^{x_2} (P_{ik}^*(x)\eta_i\eta_k + 2Q_{ik}^*(x)\eta_i\eta_k' + R_{ik}^*(x)\eta_i'\eta_k')\,dx$$
$$= \int_{x_1}^{x_2} 2\Omega^*(x, \eta, \eta')\,dx, \tag{1.5}$$

the coefficients P_{ik}^*, Q_{ik}^*, R_{ik}^*, with $P_{ik}^* = P_{ki}^*$, $R_{ik}^* = R_{ki}^*$, being of class C' on x_1x_2.

We suppose

$$\Omega(x, u, v) \leqq \Omega^*(x, u, v), \tag{1.6}$$

to hold identically in u_i and v_i for $x_1 \leqq x \leqq x_2$.

Then the conjugate points on x_1x_2 with respect to $I(\eta)$ are closer together than the conjugate points with respect to $I^*(\eta)$. More precisely: Suppose ξ^* on x_1x_2 to be conjugate to x_1 with respect to $I^*(\eta)$. Then there exists a point ξ conjugate to x_1 with respect to $I(\eta)$ with

$$x_1 < \xi \leqq \xi^*. \tag{1.7}$$

This theorem is readily proved. By definition of ξ^*, there is a set $u_i = u_i(x)$ of class C' on $x_1\xi^*$, with $u_i(x_1) = u_i(\xi^*) = 0$ and satisfying the system

$$\frac{d}{dx}\Omega_{u_i'}^* - \Omega_{u_i}^* = 0. \tag{1.8}$$

An integration by parts and (1.8) give

$$I^*(u) = \int_{x_1}^{\xi^*} 2\Omega^*(x, u, u')\,dx = \int_{x_1}^{\xi^*} (u_i\Omega_{u_i}^* + u_i'\Omega_{u_i'}^*)\,dx$$
$$= \int_{x_1}^{\xi^*} u_i\left(\Omega_{u_i}^* - \frac{d}{dx}\Omega_{u_i'}^*\right)dx = 0,$$

and this result and (1.6) give

$$I(u) \leqq I^*(u) = 0. \tag{1.9}$$

Hence $x_1\xi^*$ certainly contains some point conjugate to x_1 with respect to I, otherwise (1.9) would contradict our Lemma 1.1 applied to I for the interval $x_1\xi^*$. Hence the theorem is proved.

2. **The second variation of the length integral of a Riemannian manifold.** Let R be an open region in the space of the variables $(x_1, x_2, \cdots, x_n)$ on which a Riemannian manifold V_n is defined by the line element

$$ds^2 = a_{\alpha\beta}\,dx_\alpha\,dx_\beta. \tag{2.1}$$

This quadratic form is supposed to have its coefficients functions of the x_α of class C''' in R, and to be positive definite within this region. Let g_{12} be a geodesic arc of V_n, interior to R, and which does not intersect itself.

T. Levi-Civita has introduced (loc. cit.) a special system of coördinates $(y_1, y_2, \cdots, y_n)$, locally cartesian along g_{12}, which will be very convenient for our purpose. This system of so called *geodesic coördinates* along g_{12} has the following properties:[10]

1. The functions $y_\alpha = y_\alpha(x_1, x_2, \cdots, x_n)$ defining the transformation, are of class C''' in a neighborhood N of g_{12}, with their functional determinant $\neq 0$ in N.
2. The line element $ds^2 = b_{\alpha\beta}\, dy_\alpha\, dy_\beta$ of V_n, expressed in the new variables, has the cartesian form and is locally euclidean in every point of g_{12}, i. e.

$$b_{\alpha\beta} = \delta_{\alpha\beta}, \qquad \frac{\partial b_{\alpha\beta}}{\partial y_\gamma} = 0 \quad \text{along} \quad g_{12}. \tag{2.2}$$

3. Along g_{12} we have $y_1 = y_2 = \cdots = y_{n-1} = 0$, and hence $y_n = \sigma$ is the length of arc of g_{12} (let $\sigma = 0$ and $\sigma = l$ at 1 and 2 respectively).

The geodesics of V_n are the extremals of the problem of minimizing the integral

$$L = \int \sqrt{b_{ik}\, y_i' y_k' + 2\, b_{in}\, y_i' + b_{nn}}\, d\sigma = \int f(\sigma, y, y')\, d\sigma, \tag{2.3}$$

the variables y_i $(i = 1, 2, \cdots, n-1)$ being functions of σ. The equations $y_i \equiv 0$ give the extremal g_{12} of this problem. The second variation of L with respect to g_{12} is given by

$$I(\eta) = \int_0^l (f_{y_i y_k}\, \eta_i\, \eta_k + 2 f_{y_i y_k'}\, \eta_i\, \eta_k' + f_{y_i' y_k'}\, \eta_i'\, \eta_k')\, d\sigma, \quad (i, k = 1, 2, \cdots, n-1),$$

where the $\eta_i = \eta_i(\sigma)$ are the variations of the y_i, and the coefficients $f_{y_i y_k}, f_{y_i y_k'}$, and $f_{y_i' y_k'}$, being computed for $y_i \equiv 0$, are functions of $y_n = \sigma$ only.

By a simple direct computation combined with (2.2), one readily finds the result: Along g_{12} we have $f_{y_i y_k} = \frac{1}{2} \cdot \partial^2 b_{nn} / \partial y_i\, \partial y_k$, $f_{y_i y_k'} = 0$, $f_{y' y_k'} = \delta_{ik}$, and hence the second variation takes the form

$$I(\eta) = \int_0^l \left(\eta_i' \eta_i' + \frac{1}{2} \frac{\partial^2 b_{nn}}{\partial y_i\, \partial y_k} \eta_i\, \eta_k \right) d\sigma. \tag{2.4}$$

Furthermore we have along g_{12}

$$\frac{1}{2} \frac{\partial^2 b_{nn}}{\partial y_i\, \partial y_k} = -(ni, nk), \tag{2.5}$$

[10] We use here only the results of the §§ 1–3 of Levi-Civita's paper.

the symbol on the right hand side being a Riemann symbol of the first kind with respect to the variables y_α. In order to prove this, we start from the formula of definition of Riemann's symbols of the second kind[11]

$$\{ni, nk\} = \frac{\partial}{\partial y_k}\{nn, i\} - \frac{\partial}{\partial y_n}\{nk, i\} + \{\alpha k, i\}\{nn, \alpha\} - \{\alpha n, i\}\{nk, \alpha\}.$$

The last terms of the right side drops off because of (2.2) and also $\partial/\partial y_n \{nk, i\} = d/dy_n \{nk, i\} - \partial/\partial y_q \{nk, i\} \cdot dy_q/dy_n = 0$ along g_{12}. One therefore has $\{ni, nk\} = \partial/\partial y_k \{nn, i\}$. Furthermore $\{nn, i\} = b^{i\alpha}[nn, \alpha] = (1/2) b^{i\alpha}(2\partial b_{n\alpha}/\partial y_n - \partial b_{nn}/\partial y_\alpha)$ and because of $\partial b^{i\alpha}/\partial y_k = 0$ along g_{12}:

$$\frac{\partial}{\partial y_k}\{nn, i\} = b^{i\alpha}\left(\frac{\partial^2 b_{n\alpha}}{\partial y_n \partial y_k} - \frac{1}{2}\frac{\partial^2 b_{nn}}{\partial y_\alpha \partial y_k}\right).$$

Again

$$\frac{\partial^2 b_{n\alpha}}{\partial y_n \partial y_k} = \frac{d}{dy_n}\frac{\partial b_{n\alpha}}{\partial y_k} = 0 \quad \text{along } g_{12}$$

and finally

$$\{ni, nk\} = \frac{\partial}{\partial y_k}\{nn, i\} = -\frac{1}{2} b^{i\alpha}\frac{\partial^2 b_{nn}}{\partial y_\alpha \partial y_k} = -\frac{1}{2}\frac{\partial^2 b_{nn}}{\partial y_i \partial y_k}$$

along g_{12}. But $(ni, nk) = b^{i\alpha}\{n\alpha, nk\} = \{ni, nk\}$ along g_{12}, and thus (2.5) is proved. From (2.4) and (2.5) we get finally

$$I(\eta) = \int_0^l [\eta_i' \eta_i' - (ni, nk)\eta_i \eta_k]\, d\sigma.^{12} \tag{2.6}$$

Levi-Civita's variational equations[13]

$$\eta_i'' + (ni, nk)\eta_k = 0, \quad (i = 1, 2, \cdots, n-1), \tag{2.7}$$

are a mere consequence, as Eulers equations for $I(\eta)$ as given in (2.6).

The quadratic form $K(\sigma, \eta) = (ni, nk)\eta_i \eta_k$ which appears in (2.6), has a remarkable geometric interpretation: Let u^α and v^α be the contravariant components of two orthogonal unit vectors at a point 3 of g_{12}. One has $u^\alpha u^\alpha = 1$, $v^\alpha v^\alpha = 1$, $u^\alpha v^\alpha = 0$, and

$$K = (\alpha\beta, \gamma\delta)\, u^\alpha v^\beta u^\gamma v^\delta \,^{14} \tag{2.8}$$

[11] See for instance: T. Levi-Civita, *The absolute differential calculus*, London 1927, Chap. VII, formula (3) on pag. 175.

[12] Compare L. Berwald, loc. cit., formula (6), and J. L. Synge, loc. cit., formula (6.21). E. Cartan's formula (1) given in his paper *Sur l'écart géodésique et quelques notions connexes*, Rend. dei Lincei, serie VI, vol. 5 (1927), pp. 609–613, is a consequence of our formula (2.6).

[13] See the first system (I') on pag. 319, Math. Annalen 97.

[14] Levi-Civita, *Absolute differential calculus*, Chap. VII, formula (31) of § 10.

is the Riemannian curvature of V_n at 3 with respect to the orientation defined by the two vectors u^α and v^α. Let now v^α be the unit vector tangent to g_{12} at the point 3. One has $v^i = 0$, $v^n = 1$, $u^n = 0$, $u^i u^i = 1$, and from (2.8) one sees, that

$$K(\sigma, u) = (ni, nk)\, u^i u^k, \quad \text{with } u^i u^i = 1, \tag{2.9}$$

is the curvature of V_n at 3 with respect to the orientation tangent to g defined by the unit vector u^α orthogonal to g_{12} at the point 3.

3. On conjugate points on a geodesic arc.[15] Jacobi's condition for the extremal arc g_{12} for the problem of the shortest distance in V_n, a problem which will be discussed in § 5, depends on the discussion of the conjugate points on g_{12}. The definition of conjugate points has been given in § 1. The actual determination of the conjugate points on g_{12} can be made in every case, using the general devices of the calculus of variations. In a very special case however, this determination is immediate, namely the case when V_n is of constant positive curvature K. From (2.9) one gets in this case $(ni, nk)\,\eta_i \eta_k = K \eta_i \eta_i$ and the system (2.7) becomes

$$\eta_i'' + K\eta_i = 0, \qquad (i = 1, 2, \cdots, n-1).$$

Its solution $\eta_i = \sin(K^{1/2}\sigma)$ and the definition of conjugate points (§ 1) give the

THEOREM 3.1. *Let g_{12} be a geodesic arc of a Riemannian manifold V_n of positive constant curvature K. The distance along g_{12} between two consecutive conjugate points on g_{12} is constant and equal to $\pi/\sqrt{K}$.*

The general device for the determination of the conjugate point of $1\,(\sigma = 0)$, is as follows.[16] Determine the set of $n-1$ solutions $u_{1k}, u_{2k}, \cdots, u_{n-1,k}$, $(k = 1, 2, \cdots, n-1)$, of the system (2.7), which is uniquely defined by the initial values

$$u_{ik}(0) = 0, \qquad u_{ik}'(0) = \delta_{ik},$$

where (δ_{ik}) could be replaced by any other matrix of values with determinant $\neq 0$, and set up the equation

$$\Delta(\sigma) \equiv |u_{ik}(\sigma)| = 0.$$

Its first positive root $\leqq l$, if there is one, will determine on g_{12} the first conjugate point of $1\,(\sigma = 0)$.

We shall apply this method in the case when $K(\sigma, u)$ defined by (2.9) is constant along g_{12}, for u^i propagated parallel to itself in the sense of

[15] J. L. Synge considers (loc. cit. § 10) only conjugate points on g_{12} which are conjugate with respect to some definite surface V_2 containing the arc g_{12}.

[16] See J. Hadamard, *Calcul des Variations*, §§ 282–284.

Levi-Civita. Since Christoffel's symbols vanish identically along g_{12}, this means that $K(\sigma, u)$ has to be constant if the u^i are constants. Hence the (ni, nk) are to be constants along g_{12}, and we put $(ni, nk) = a_{ik}$, with $a_{ik} = a_{ki}$. The system (2.7) or $\eta_i'' + a_{ik}\eta_k = 0$ represents a linear transformation with a real symmetric matrix of coefficients and a well known theorem[17] insures the existence of a non-singular transformation $\eta_i = c_{ik}\bar{\eta}_k$ with constant coefficients, carrying (2.7) into its canonical form

$$\bar{\eta}_i'' + \mu_{(i)}\bar{\eta}_i = 0, \qquad (i = 1, 2, \cdots, n-1). \tag{3.1}$$

In these equations, the μ_i are the $n-1$ roots of the characteristic equation

$$|a_{ik} - \mu\delta_{ik}| = 0, \tag{3.2}$$

and the subscript in parantheses of $\mu_{(i)}$ will indicate that there is no summation on the left hand side of (3.1). Let $\mu_r > 0$ for $r = 1, 2, \cdots, r'$, $\mu_s < 0$ for $s = r'+1, \cdots, r'+s'$, and $\mu_t = 0$ for $t = r'+s'+1, \cdots, r'+s'+t' = n-1$. We set up a set of $n-1$ solutions of the system (3.1) as follows: We put $\bar{u}_{ik} = 0$ for $i \neq k$, and for $i = k$ we put $\bar{u}_{kk} = \mu_k^{-1/2} \sin(\mu_k^{1/2}\sigma)$, $2^{-1}|\mu_k|^{-1/2}(e^{|\mu_k|^{1/2}\sigma} - e^{-|\mu_k|^{1/2}\sigma})$, or σ, according as $\mu_k > 0$, $\mu_k < 0$, or $\mu_k = 0$. Returning to the original system (2.7), we get

$$\Delta(\sigma) = |u_{ik}| = |c_{ik}| \cdot |\bar{u}_{ik}| = \varphi(\sigma) \prod_{r=1}^{r'} \sin(\mu_r^{1/2}\sigma),$$

the function $\varphi(\sigma)$ being $\neq 0$ for $\sigma > 0$. This last formula proves the following

THEOREM 3.2. *Let g_{12} be a geodesic arc of a Riemannian manifold V_n, along which the curvature K is constant with respect to the orientations tangent to g_{12} and defined by a unit vector orthogonal to g_{12} which is moved along g_{12} parallel to itself in the sense of Levi-Civita. As a consequence, the Riemann symbols (ni, nk), $(i, k = 1, 2, \cdots, n-1)$, of the manifold V_n with respect to its geodesic coördinates along g_{12}, are constant along g_{12}. Let μ be the largest root of the characteristic equation*

$$|(ni, nk) - \mu\delta_{ik}| = 0.$$

For $\mu \leqq 0$, there are no conjugate points on g_{12}. If $\mu > 0$, the distance along g_{12} between two consecutive conjugate points of g_{12} is constant and equal to $\pi/\sqrt{\mu}$.

We shall prove now the following

[17] See H. Hilton, *Homogeneous linear Substitutions*, Oxford 1914, Chap. I, § 12, pag. 43.

THEOREM 3.3. *Let g_{12} be a geodesic arc of two Riemannian manifolds V_n and V_n^* which osculate*[18] *each other along g_{12}. We suppose that we have for their curvatures the relation*

$$K^* \leqq K, \tag{3.3}$$

for every orientation tangent to g_{12}. Then the conjugate points on g_{12} with respect to V_n are closer together than the conjugate points with respect to V_n^* (in the same sense as in Theorem 1.1).

Let $ds^{*2} = a^*_{\alpha\beta}\, dx_\alpha\, dx_\beta$ be the line element of V_n^*. The osculation is expressed by the relation $a^*_{\alpha\beta} = a_{\alpha\beta}$, $\partial a^*_{\alpha\beta}/\partial x_\gamma = \partial a_{\alpha\beta}/\partial x_\gamma$ along g_{12}. Any transformation of coördinates preserves these relations and therefore has $ds^{*2} = b^*_{\alpha\beta}\, dy_\alpha\, dy_\beta$ the same properties (2.2) as ds^2, i. e. the variables y_i are geodesic coördinates along g_{12} also for the manifold V_n^*. For V_n^*, (2.6) becomes

$$I^*(\eta) = \int_0^l [\eta_i'\, \eta_i' - (ni, nk)^*\, \eta_i\, \eta_k]\, d\sigma. \tag{3.4}$$

The inequality (3.3), the expression (2.9) and the similar expression for V_n^* show, that the inequality (1.6) is verified for our integrals I and I^* as given by (2.6) and (3.4). Hence Theorem 1.1 implies Theorem 3.3.

Let V_n^* osculate V_n along g_{12}. One knows that the curvature K^* of V_n^* is expressible, for orientations tangent to g_{12}, as a quadratic form like (2.9) in the components u^i defining the orientation. Of some interest seems to be the following inverse question: *Let $p_{ik}(\sigma) u^i u^k$ be a given quadratic form along g_{12}. Is there a manifold V_n^* osculating V_n along g_{12} and whose curvature along g_{12} is given by $K^*(\sigma, u) = p_{ik}(\sigma) u^i u^k$, $(u^i u^i = 1)$?*

The solution is obvious: *The manifold V_n^* whose metric is defined by $ds^{*2} = dy_1^2 + \cdots + dy_{n-1}^2 + (1 - p_{ik}(\sigma)\, y_i y_k)\, d\sigma^2$, $(d\sigma = dy_n)$, in some neighborhood of g_{12}, satisfies the proposed condition.* Indeed, this manifold V_n^* certainly osculates V_n along g_{12}, and (2.9) and (2.5), applied to V_n^* give $K^*(\sigma, u) = (ni, nk)^*\, u^i u^k = p_{ik}(\sigma)\, u^i u^k$.

As another indirect corollary of Theorem 1.1, I mention the following

THEOREM 3.4. *On a geodesic arc g_{12} along which the curvature K of V_n takes on no positive value with respect to the orientations tangent to g_{12}, there are never conjugate points.*

Its proof is as follows: We set up the integrals $I(\eta)$ and $I^*(\eta)$ of Theorem 1.1 with $2\Omega = \eta_i'\, \eta_i'$ and $2\Omega^* = \eta_i'\, \eta_i' - (ni, nk)\, \eta_i\, \eta_k$. The condition $K \leqq 0$ along g_{12} and formula (2.9) show, that the inequality (1.6) holds identically. The existence of a point on g_{12} conjugate to 1 with

[18] See E. Cartan, *La géométrie des espaces de Riemann*, Paris 1928, Chap. IV.

respect to $I^*(\eta)$, would imply the existence of a similar point with respect to the integral

$$I(\eta) = \int_0^l \eta_i' \eta_i' \, d\sigma.$$

Its Euler equations being $\eta_i'' = 0$ $(i = 1, 2, \cdots, n-1)$, this last integral has no conjugate points. This proves the Theorem 3.4.

4. **A generalization of O. Bonnet's theorem on ovaloid surfaces.** On the Riemannian manifold V_n of § 2, we make the following additional assumptions:

(a) There is a positive constant A such that

$$K \geqq 1/A^2 \tag{4.1}$$

holds throughout the open region R on which V_n is defined.

(b) Any two points 1 and 2 of R can be joined by an arc g_{12} interior to R and whose length l_{12} is the absolute shortest distance between 1 and 2 within V_n, which implies that g_{12} is a smooth geodesic arc.

We shall prove the following

THEOREM 4.1. *The assumptions a and b imply the inequality*

$$l_{12} \leqq \pi A, \tag{4.2}$$

l_{12} being the length of the shortest geodesic arc g_{12} joining two points 1 *and* 2 *of R and whose existence is insured by assumption b.*

Let 1 and 2 be any two points of V_n and g_{12} be the geodesic arc given by assumption (b). Let $I(\eta)$, given by (2.6), be the second variation of the length integral for this arc g_{12}. In order to prove (4.2), let us suppose the contrary:

$$l_{12} > \pi A. \tag{4.3}$$

We set up the integral

$$I^*(\eta) = \int_0^{l_{12}} \left(\eta_i' \eta_i' - \frac{1}{A^2} \eta_i \eta_i \right) d\sigma. \tag{4.4}$$

From (2.9) and (4.1) we get

$$\eta_i' \eta_i' - (ni, nk) \eta_i \eta_k \leqq \eta_i' \eta_i' - \frac{1}{A^2} \eta_i \eta_i$$

for $0 \leqq \sigma \leqq l_{12}$, η_i and η_i' arbitrary. The points $\sigma = 0$ and $\sigma = \pi A$ are conjugate with respect to $I^*(\eta)$ and Theorem 1.1 shows that there is a point on g_{12}, conjugate to 1 $(\sigma = 0)$ with respect to $I(\eta)$ and which lies between $\sigma = 0$ and $\sigma = \pi A$. Our assumption (4.3) shows that this point is actually interior to g_{12}. This last result contradicts the minimum property of g_{12}, Jacobi's condition being necessary for the existence of a relative minimum. Hence (4.3) is impossible and the theorem is proved.

The portion of the surface of a sphere in S_3 interior to some small circle of it, satisfies the assumptions (a) and (b), but the remaining portion, exterior to this circle does not. Therefore the following extension of Theorem 4.1 seems to be of interest:

COROLLARY 4.1. *We retain assumption* (a), *but replace assumption* (b) *by the following assumption* (b'): *Any two points* 1 *and* 2 *of* V_n *can be joined by a broken curve composed of* m *smooth geodesic arcs* $g_{11'}, g_{1'2'}, \cdots, g_{m-1',2}$, *every one of these arcs furnishing the absolute shortest distance in* V_n *between its two endpoints. There is then a path from* 1 *to* 2 *within* V_n *whose length is* $\leqq m\pi A$.

Theorem 4.1 gives indeed $g_{11'} \leqq \pi A$, $g_{1'2'} \leqq \pi A$, $\cdots$, $g_{m-1',2} \leqq \pi A$, and therefore $g_{11'} + g_{1'2'} + \cdots + g_{m-1',2} \leqq m\pi A$. That this limit $m\pi A$ which, according to the Corollary 4.1, is an upper bound for the greatest lower bound of the distances in V_n between any two points of V_n, can be approached as nearly as may be desired for some pair of points in a suitable choosen manifold, can be seen already for $n = 2$ using a sphere suitably cut: We cut the sphere with radius A along its equator, leaving both hemispheres connected only by a small segment of this equator. This manifold V_2 satisfies both conditions (a) and (b') with $m = 2$. If we take 1 and 2 in different hemispheres, close to the equator and almost diametrally opposite to the connecting segment of the equator, then the shortest way joining 1 and 2, will be as close to $2\pi A$ as we want, choosing the connecting segment sufficiently small. A shere suitably cut along $m-1$ circles proves the same fact for any value of m.

5. The problem of the shortest distance in V_n.[19] Let g_{12} be a geodesic arc in the manifold V_n of § 2. *What are the conditions which insure that* g_{12} *represents the shortest curve joining 1 and 2, compared with all the other curves* C_{12} *of a sufficiently small neighborhood of* g_{12} *and joining these two points?* We have a so called parametric problem of the calculus of variations, the problem of minimizing the integral

$$L = \int_{t_1}^{t_2} \sqrt{a_{\alpha\beta}\, x'_\alpha x'_\beta}\, dt = \int_{t_1}^{t_2} f(x, x')\, dt, \tag{5.1}$$

$x_\alpha = x_\alpha(t)$, $(t_1 \leqq t \leqq t_2)$, being a parametric representation of C_{12}. We suppose the functions $x_\alpha = x_\alpha(t)$ to be of class D' and $x'_\alpha x'_\alpha > 0$ for $t_1 \leqq t \leqq t_2$, with $x_\alpha(t_1) = x_{\alpha 1}$ and $x_\alpha(t_2) = x_{\alpha 2}$, the whole curve C_{12} lieing within some neighborhood of g_{12}, interior to the open region R of § 2.

The arc g_{12} was supposed to be a geodesic arc, and it is well known that this is a necessary condition for the minimum property of g_{12}, since

[19] For the same problem on a surface ($n = 2$) see G. Darboux, *Théorie générale des surfaces*, vol. III, Paris 1894, Chap. V.

the Weierstrass-Erdmann corner conditions remove the possibility of corners for g_{12}. The Weierstrass function[20]

$$E(x, x', p) = f(x, p) - f(x, x') - (p_\alpha - x'_\alpha)\, f_{x'_\alpha}(x, x'),$$

takes in our case the form

$$(5.2)\quad E(x, x', p) = (a_{\alpha\beta}\, x'_\alpha\, x'_\beta)^{-1/2}\, [(a_{\alpha\beta}\, x'_\alpha\, x'_\beta)^{1/2}\, (a_{\alpha\beta}\, p_\alpha\, p_\beta)^{1/2} - a_{\alpha\beta}\, x'_\alpha\, p_\beta].$$

Legendre's condition is related to the quadratic form

$$(5.3)\quad \begin{aligned} Q(x, x', p, p) &= f_{x'_\alpha x'_\beta}\, p_\alpha\, p_\beta \\ &= (a_{\alpha\beta} x'_\alpha x'_\beta)^{-3/2}\, [(a_{\alpha\beta} p_\alpha p_\beta)\, (a_{\alpha\beta} x'_\alpha x'_\beta) - (a_{\alpha\beta} p_\alpha x'_\beta)^2]. \end{aligned}$$

Let $x_\alpha = x_\alpha(t)$, $(t_1 \leqq t \leqq t_2)$, be a parametric representation of g_{12}. From (5.2) and (5.3) one concludes that $E(x, x', p) > 0$ and $Q(x, x', p, p) > 0$, at any point of g_{12} and for every set p_α with $p_\alpha\, p_\alpha > 0$ and which is not proportional to the set x'_α: *The necessary conditions of Weierstrass and Legendre are always satisfied along g_{12} in the strict sense.*[21]

A sufficiency theorem of Bliss[22] for the general parametric problem leads to the following result:

THEOREM 5.1. *Let g_{12} be a geodesic arc of V_n, interior to V_n and which does not intersect itself. If g_{12} does contain no conjugate point to* 1, *then g_{12} actually is the shortest curve joining* 1 *and* 2, *compaired with all such curves of class D' of a sufficiently small neighborhood of g_{12}.*[23]

Theorems 5.1 and 3.1 show, that a geodesic arc g_{12} of a manifold of constant positive curvature K, gives a relative shortest distance as far as its length does not reach the value $\pi/\sqrt{K}$. Theorems 5.1 and 3.4 show, that a geodesic arc g_{12} of a manifold of non-positive curvature, which does not intersect itself, always gives a relative shortest distance between its endpoints.[24]

---

[20] See G. A. Bliss, *The Weierstrass E-function for problems of the calculus of variations in parametric form*, Transactions of the American Math. Society, vol. 15 (1914), pp. 369–378.

[21] Compare A. R. Forsyth, *Geometry of four dimensions*, vol. II, Cambridge 1930, Chap. XVII, § 299, where the same fact is proved for a V_2 which is emmersed in a euclidean space S_4.

[22] This theorem, which Professor Bliss uses to give in his lectures, insures the existence of a relative minimum for the parametric problem in n-space for which the Legendre and Weierstrass necessary conditions are satisfied in the strict sense.

[23] A. R. Forsyth gets, loc. cit., the same result for V_2, based however on no sufficiency theorem.

[24] Compare the note III at the end of E. Cartan's *Géométrie des espaces de Riemann*, Paris 1928, where this result is obtained by a quite different method.

UNIVERSITY OF CHICAGO, CHICAGO, ILLINOIS.

GROUPS OF MOTIONS IN SPACES ADMITTING ABSOLUTE PARALLELISM.[1]

By H. P. Robertson.

I. Introduction.

One of the most fundamental questions which can be asked concerning a space of given structure is whether it admits *automorphisms* or not, i. e. whether it is possible to establish one or more correspondences $P \rightarrow \bar{P}$ of the space on itself in such a way that corresponding points P, $\bar{P}$ or configurations of such points are *intrinsically*—by means of the structure of the space alone—indistinguishable. Obviously the set of all such automorphisms, including the identity $P \rightarrow P$, admitted by such a space constitutes a *group*; in case this group involves r essential parameters σ_u $(u = 1, 2, \cdots, r)$ we say that the space admits an *r-parameter group of motions* into itself. The converse of this problem, that of determining the restrictions placed on the structure of a space by *a priori* symmetry conditions, is not only of geometrical interest, but is further of importance to any physical theory which considers the structure of physical space or space-time as contingent on its physical content—such as the state and distribution of matter and electricity.

The discussion of such groups of motions in spaces defined by a Riemannian metric has been given by Killing, Fubini and other writers[2]. Einstein's 1916 theory of relativity, which is based on the assumption that physical space-time constitutes a 4-dimensional Riemannian manifold, leads to applications of these investigations to problems met in physics and astronomy. His more recent unified field theory[3], in which the Riemannian space-time is assumed to admit absolute parallelism, raises the corresponding problems for such a space; two particular investigations of this type have been carried out by Einstein and Mayer[4] and by McVittie[5].

It is the purpose of the present paper to develop the theory of groups of automorphisms in general absolute parallelism spaces, with special

[1] Received December 4, 1931.

[2] For a concise development of this theory and for references to the literature see L. P. Eisenhart "Riemannian Geometry", Chap. VI, pp. 221–251 (Princeton 1926); hereinafter referred to as "R. G.".

[3] A. Einstein, various papers in the Preuß. Akad. Sitzungsber. 1928–1931.

[4] A. Einstein and W. Mayer „Zwei strenge statische Lösungen der Feldgleichungen der einheitlichen Feldtheorie", Preuß. Akad. Sitzungsber. 1930, pp. 110–120.

[5] G. C. McVittie "Solutions with Axial Symmetry of Einstein's Equations of Teleparallelism", Edinburgh Math. Soc. Proc. (2) **2**, pp. 140–150 (1930).

reference to Riemannian spaces admitting an absolute parallelism. The determination, in principle, of the complete group admitted by a space of given structure is given, and the spaces admitting certain groups of interest to the general theory are obtained. The results are finally applied to certain Riemannian absolute parallelism spaces ("R. A. P. S.") of possible physical interest—including that considered by Einstein and Mayer—but in this we restrict ourselves entirely to geometrical considerations, leaving entirely aside the problem of the further specification of space-time by the introduction of field equations.

We first discuss those properties of general and of Riemannian absolute parallelism spaces which will be required in the sequel.

1. **Notation and general preliminaries.** The n-space T_n with which we are concerned is characterized by the property of *absolute parallelism*: with each vector λ_0 at a point P_0 of T_n is associated a unique vector $\lambda(P)$ at an arbitrary point P of T_n, which is said to be *parallel* to $\lambda_0 : \lambda(P) \parallel \lambda_0$. This distant parallelism is a reflexive, symmetric and transitive relation such that from $\lambda(P) \parallel \lambda_0$, $\mu(P) \parallel \mu_0$ we may conclude that $a\lambda(P) \parallel a\lambda_0$, $\lambda(P)+\mu(P) \parallel \lambda_0+\mu_0$. For our present purposes we may assume that on introducing a coördinate system x^μ ($\mu = 1, 2, \cdots, n$) the components $\lambda^\mu(x)$ of the vector $\lambda(P)$ are analytic functions of the x^μ—although it is only necessary to assume that they possess derivatives of a certain order.

Let $\lambda_i^\mu(x)$ ($i = 1, 2, \cdots, n$) be the n vectors at P which are associated with n independent vectors at some given point P_0; this ennuple of n (independent) vectors then defines the structure of the space. Any other ennuple $\bar{\lambda}_i^\mu(x)$ which may be used to define the structure can be obtained from the vectors $\lambda_i^\mu(x)$ by a non-singular linear homogeneous transformation with constant coefficients:

$$\bar{\lambda}_i^\mu(x) = a_i^j \lambda_j^\mu(x), \tag{1.1}$$

for the vectors $\bar{\lambda}_i^\mu(x_0)$ at P_0 which are parallel to the $\bar{\lambda}_i^\mu(x)$ at P must be independent linear combinations of the original $\lambda_i^\mu(x_0)$ with numerical coefficients a_i^j. We may associate with the ennuple $\lambda_i^\mu(x)$ of contravariant vectors the reciprocal ennuple $\lambda_\mu^i(x)$ of covariant vectors defined by the equations

$$\lambda_i^\mu \lambda_\nu^i = \delta_\nu^\mu \quad \text{or} \quad \lambda_\mu^i \lambda_j^\mu = \delta_j^i; \tag{1.2}$$

these covariant vectors each undergo the transformation contragredient to (1.1) on transition to the new ennuple $\bar{\lambda}_i^\mu$. Any tensor $A_{\nu\cdots}^{\mu\cdots}$ may be expressed in terms of its scalar components $A_{j\cdots}^{i\cdots}$ with respect to the ennuple:

$$A_{\nu\cdots}^{\mu\cdots} = A_{j\cdots}^{i\cdots} \lambda_i^\mu \cdots \lambda_\nu^j, \quad \text{where} \quad A_{j\cdots}^{i\cdots} \equiv A_{\nu\cdots}^{\mu\cdots} \lambda_\mu^i \cdots \lambda_j^\nu. \tag{1.3}$$

On change of ennuple the scalar components $A^{i\cdots}_{j\cdots}$ transform as tensors of indicated rank under the affine transformation defined by (1.1) for a covariant vector. Similarly a pure tensor $A^{\mu\cdots}_{\nu\cdots}$, with no free Latin indices, is an invariant under (1.1).

The association with a vector at $P(x)$ of a parallel vector at $\bar{P}(x+dx)$ may be expressed in terms of a parallel displacement with coefficients of connection[6]

$$L^{\mu}_{\nu\varrho} = -\lambda^{i}_{\nu}\frac{\partial \lambda^{\mu}_{i}}{\partial x^{\varrho}} \left(= \frac{\partial \lambda^{i}_{\nu}}{\partial x^{\varrho}}\lambda^{\mu}_{i}\right). \tag{1.4}$$

Covariant differentiation with respect to this connection will be indicated by a comma; naturally (1.4) is exactly the condition that the covariant derivatives $\lambda^{\mu}_{i,\nu}$, $\lambda^{i}_{\mu,\nu}$ of the ennuple vectors vanish. The anti-symmetric part

$$\Omega^{\mu}_{\nu\varrho} = L^{\mu}_{[\nu\varrho]} = \frac{1}{2}(L^{\mu}_{\nu\varrho} - L^{\mu}_{\varrho\nu}) \tag{1.5}$$

of $L^{\mu}_{\nu\varrho}$ satisfies the identities[7]

$$\Omega^{\mu}_{[\nu\varrho,\sigma]} + 2\,\Omega^{\tau}_{[\nu\varrho}\,\Omega^{\mu}_{\sigma]\tau} = 0. \tag{1.6}$$

The scalar invariant associated with the covariant derivative $A^{\mu\cdots}_{\nu\cdots,\varrho}$ of the tensor $A^{\mu\cdots}_{\nu\cdots}$ is the intrinsic derivative

$$A^{i\cdots}_{j\cdots,k} \equiv \frac{\partial A^{i\cdots}_{j\cdots}}{\partial s^{k}} = \lambda^{\varrho}_{k}\frac{\partial A^{i\cdots}_{j\cdots}}{\partial x^{\varrho}} \tag{1.7}$$

of the scalar components $A^{i\cdots}_{j\cdots}$. The Ricci identity, in terms of scalars, is

$$A^{i\cdots}_{j\cdots,[k,l]} = -A^{i\cdots}_{j\cdots,m}\,\Omega^{m}_{kl}, \tag{1.8}$$

where the Ω^{m}_{kl} are the scalar components of the tensor $\Omega^{\mu}_{\nu\varrho}$.

We shall have occasion to refer to the concept of an *autoparallel subspace* T_{n-m} *of* T_n, i. e. an $(n-m)$-dimensional variety immersed in T_n which is such that any vector lying in T_{n-m} remains wholly in it under an arbitrary parallel displacement in the sense of T_n. To find the necessary and sufficient conditions that a given T_{n-m} be autoparallel we introduce a coördinate system x^{μ}: x^{α}, x^{ω} ($\alpha = 1, 2, \cdots, m$; $\omega = m+1, m+2, \cdots, n$) in which the equations of T_{n-m} are $x^{\alpha} = 0$ and in which a given point P_0 of T_{n-m} has the coördinates $x^{\omega} = 0$. The vector λ^{μ} at P_0 then lies in T_{n-m} provided

---

[6] Cf. L. P. Eisenhart "Non-Riemannian Geometry", p. 48 (New York 1927); hereinafter referred to as "N.-R. G." The notation employed in the present paper follows, on the whole, that of this text and R. G.

[7] A. Einstein "Zur einheitlichen Feldtheorie", Preuß. Akad. Sitzungsber. 1929, eq. (3) p. 4. The brackets around the indices $\nu\varrho\sigma$ indicate 1/6 the alternating sum with respect to these indices; cf. J. A. Schouten "Der Ricci-Kalkül", p. 4 (Berlin 1924).

$$\lambda^\alpha = \lambda^i \lambda_i^\alpha(0) = 0, \tag{1.9}$$

and the vector $\lambda^\mu(x^\omega)$ at an arbitrary point $P(x^\omega)$ of T_{n-m} which is parallel to λ^μ at P_0 lies wholly in T_{n-m} provided

$$\lambda^\alpha(x^\omega) \equiv \lambda^i \lambda_i^\alpha(x^\omega) = 0 \tag{1.10}$$

where the λ^i are the same constants as in (1.9). The condition that T_{n-m} be autoparallel is that the defining ennuple be such that (1.10) must be satisfied for any constants λ^i satisfying (1.9). To obtain an explicit form of these conditions we transform to a new ennuple such that at P_0 $\lambda_i^\mu(0) = \delta_i^\mu$; eqs. (1.9) then become $\lambda^a = 0$ $(a = 1, 2, \cdots, m)$ and (1.10) must therefore be satisfied for arbitrary λ^i of the form $\lambda^a = 0$, λ^z $(z = m+1, m+2, \cdots, n)$. Hence the necessary and sufficient condition that T_{n-m} be autoparallel is that it be possible to choose a coördinate system and an ennuple in such a manner that

$$\lambda_z^\alpha(x^\omega) = 0 \quad \begin{pmatrix} \alpha = 1, 2, \cdots, m \\ z, \omega = m+1, m+2, \cdots, n \end{pmatrix} \tag{1.11}$$

at all points x^μ: $x^\alpha = 0$, x^ω of T_{n-m}.

Riemannian space with absolute parallelism. A case of particular interest arises when we consider the ennuple defining the parallelism as an orthogonal ennuple of unit vectors in a Riemannian space—the resulting geometry, for $n = 4$, is that employed by Einstein in his recent unified field theory. We then have in addition to the ennuple the Riemannian metric defined by the following conditions: the square of the absolute magnitude of the ith vector shall be e_i $(= \pm 1)$, and any two distinct vectors of the ennuple shall be orthogonal. These conditions are, however, equivalent to the assertion that the scalar components of the metric tensor $g_{\mu\nu}$ are

$$g_{ij} \equiv g_{\mu\nu} \lambda_i^\mu \lambda_j^\nu = e_i \delta_{ij}, \tag{1.12}$$

where the indices appearing on the e_i are to be disregarded in applying the summation convention to Latin indices. The $g_{\mu\nu}$ are themselves given by

$$g_{\mu\nu} = g_{ij} \lambda_\mu^i \lambda_\nu^j = \lambda_\mu^i \lambda_{i\nu}; \tag{1.13}$$

the signature of the Riemannian space which they define is equal to the difference between the number of e_i which are $+1$ and the number which are -1. We have here introduced the convention that a Latin index may be raised or lowered in the usual manner by means of the form (1.12); that this convention is permissable is due to the fact that we must now restrict the transformations (1.1) to a new ennuple to those which leave the form defined by (1.12) invariant, in order that the metric (1.13) be

invariant under change of ennuple. These transformations, which we refer to as *rotations,* must satisfy the conditions

$$a_i^{\cdot k}\, a_j^{\cdot l}\, g_{kl} = g_{ij}, \quad \text{or} \quad a^{ik}\, a_{jk} = \delta_j^i. \tag{1.14}$$

An infinitesimal rotation is defined by

$$a_i^{\cdot j} = \delta_i^j + \alpha_i^{\cdot j} + \cdots \tag{1.15}$$

where the first order terms satisfy the conditions

$$2\,\alpha_{(ij)} = \alpha_{ij} + \alpha_{ji} = 0. \tag{1.16}$$

Greek indices may now be raised or lowered with the aid of the $g_{\mu\nu}$, and we have in particular

$$\lambda_\mu^i = g^{ij}\, g_{\mu\nu}\, \lambda_j^\nu = g_{\mu\nu}\, \lambda^{i\nu}. \tag{1.17}$$

It is to be observed that although in a general absolute parallelism space the ennuple may always be chosen in such a manner that at a given point P_0 it assumes the form $\lambda_i^\mu = \delta_i^\mu$ without restriction on the coördinate system, in a Riemannian absolute parallelism space this choice of ennuple at P_0 is possible if and only if the coördinate system is chosen in such a way that the $g_{\mu\nu}$ assume the canonical form g_{ij} at P_0.

2. **Equivalence conditions.** In order that the two absolute parallelism spaces $\lambda_i^\mu(x)$, $\overline{\lambda}_i^\mu(\overline{x})$ be equivalent there must exist a correspondence $x^\mu \to \overline{x}^\mu$ and an affine transformation (1.1) such that

$$\overline{\lambda}_i^\mu(\overline{x}) = a_i^{\cdot j}\, \frac{\partial\, \overline{x}^\mu}{\partial\, x^\nu}\, \lambda_j^\nu(x) \tag{2.1}$$

or

$$\frac{\partial\, \overline{x}^\mu}{\partial\, x^\nu} = \overline{\lambda}_{\bar{i}}^\mu(\overline{x})\, a_{\cdot i}^{\bar{i}}\, \lambda_\nu^i(x), \tag{2.2}$$

where the $a_{\cdot i}^{\bar{i}}$ define the inverse of the transformation (1.1). The conditions of integrability of these equations are[8]

$$\overline{\Omega}_{\bar{j}\,\bar{k}}^{\bar{i}}(\overline{x}) = a_{\cdot i}^{\bar{i}}\, a_{\bar{j}}^{\cdot j}\, a_{\bar{k}}^{\cdot k}\, \Omega_{jk}^i(x), \tag{2.3}$$

$$\begin{aligned}
\overline{\Omega}_{\bar{j}\,\bar{k},\bar{l}}^{\bar{i}}(\overline{x}) &= a_{\cdot i}^{\bar{i}}\, a_{\bar{j}}^{\cdot j}\, a_{\bar{k}}^{\cdot k}\, a_{\bar{l}}^{\cdot l}\, \Omega_{jk,l}^i(x), \\
&\vdots \qquad\qquad \vdots \\
\overline{\Omega}_{\bar{j}\,\bar{k},\bar{l}_1\cdots\bar{l}_\alpha}^{\bar{i}}(\overline{x}) &= a_{\cdot i}^{\bar{i}}\, a_{\bar{j}}^{\cdot j} \cdots a_{\bar{l}_\alpha}^{\cdot l_\alpha}\, \Omega_{jk,l_1\cdots l_\alpha}^i(x), \\
&\vdots \qquad\qquad \vdots
\end{aligned} \tag{2.4}$$

[8] For a modern account of the theory of such equations and for references see N.-R. G., p. 14.

i. e. the scalars $\bar{\Omega}$ and their intrinsic derivatives must be equal to the scalars Ω and their corresponding derivatives to within an affine transformation. The precise criterion for equivalence may now be stated in the following form: *the two spaces* $\lambda_i^\mu(x)$, $\bar{\lambda}_i^\mu(\bar{x})$ *are equivalent provided there exists an affine transformation* a_i^j *and a non-negative integer* N *such that* 1. *the first* N *sets of* (2.3), (2.4) *are compatible equations for the variables* $\bar{x}$ *as functions of the independent variables* x, *and* 2. *all solutions of these equations satisfy the* $(N+1)$*st set.*

In the special case in which the first set (2.3) of conditions is satisfied identically in the $\bar{x}$, x for some choice of the a_i^j, equations (2.1) are completely integrable and the complete solutions contains, in addition to the a_i^j, n arbitrary constants (e. g. the values of the $\bar{x}^\mu$ for $x^\mu = 0$).

Riemannian absolute parallelism space. The above considerations hold for a Riemannian space admitting absolute parallelism provided we require that the constants a_i^j define a rotation (1.14).

II. General Theory of Groups of Motions. The Group Admitted by a Given Space.

In this section the fundamental equations for a group of motions in an absolute parallelism space are developed and the results are applied to the problem of determining the complete group of motions admitted by a space of given structure.

3. **Fundamental equations.** We now return to the problem of finding the conditions that a T_n admit an automorphism in the sense discussed in the introduction. But to say that T_n admits such an isomorphism is the same as saying that it admits a non-trivial equivalence with itself in the sense of § 2; the fundamental equations for an automorphism are therefore obtained from (2.1) on replacing $\bar{\lambda}_i^\mu(\bar{x})$ by $\lambda_i^\mu(\bar{x})$:

$$\lambda_i^\mu(\bar{x}) = a_i^j \frac{\partial \bar{x}^\mu}{\partial x^\nu} \lambda_j^\nu(x). \tag{3.1}$$

The conditions of integrability of these equations are given by (2.3), (2.4) on replacing the scalars $\bar{\Omega}(\bar{x})$ and their intrinsic derivatives by the $\Omega(\bar{x})$ and their corresponding derivatives.

As remarked, the set G of all such automorphisms A: $x^\mu \to x^\mu$, a_i^j admitted by a given space, including the identity $\mathbf{1}$: $x^\mu \to x^\mu$, δ_i^j, constitutes a group. For a formal proof we need only remark that G contains 1) the

identity **1**, 2) the inverse A^{-1}: $\bar{x}^\mu \to x^\mu$, $a_i^{-1j} = a^j_{\cdot i}$ of any A, and 3) the product $C = BA$: $x^\mu \to \tilde{x}^\mu$, $c_i^{\cdot j}$ of any two automorphisms A and B: $\bar{x}^\mu \to \tilde{x}^\mu$, $b_i^{\cdot j}$, where

$$c_i^{\cdot j} = b_i^{\cdot k}\, a_k^{\cdot j}. \tag{3.2}$$

Denoting the group of coördinate transformations $x^\mu \to \bar{x}^\mu$ associated with G by $G(x)$ and the corresponding isomorphic group of affine transformations $a_i^{\cdot j}$ by $G(\lambda)$, the conditions of composition (3.2) assert that the group $G(\lambda)$ is an n-dimensional representation of the transformation group $G(x)$. Now those automorphisms for which $a_i^{\cdot j}$ is the identity δ_i^j, which we shall in the following call *special*, are associated with an invariant sub-group $g(x)$ of $G(x)$—for the conjugate $A' = BAB^{-1}$ of a special automorphism A with respect to an arbitrary automorphism B is itself special. The affine representation $G(\lambda)$ of $G(x)$ is therefore a faithful representation of the factor group $G(x)/g(x)$.[9]

In case equations (3.1) admit a family of solutions

$$A(\sigma)\colon \quad x^\mu \to \bar{x}^\mu = f^\mu(x, \sigma), \qquad a_i^{\cdot j} = a_i^{\cdot j}(\sigma) \tag{3.3}$$

depending on r essential continuous parameters σ_u $(u = 1, 2, \cdots, r)$ which for some value of the σ_u—say $\sigma_u = 0$—yields the identity, then by the above these automorphisms constitute a *continuous group* G_r: $G_r(x)$, $G_r(\lambda)$ *of motions* of the space into itself. We shall, as is usual in considerations involving continuous groups of transformations, concern ourselves with the infinitesimal elements

$$\delta\sigma\colon \quad \delta x^\mu = \xi^\mu(x) \cdot \delta\sigma, \qquad \delta a_i^{\cdot j} = \alpha_i^{\cdot j} \cdot \delta\sigma \tag{3.4}$$

of the group. Equations (3.1) then become

$$[\xi\, \lambda_i]^\mu \equiv \xi^\nu \frac{\partial \lambda_i^\mu}{\partial x^\nu} - \lambda_i^\nu \frac{\partial \xi^\mu}{\partial x^\nu} = \alpha_i^{\cdot j}\, \lambda_j^\mu; \tag{3.5}$$

alternative forms of these equations, which will be useful in the sequel, are

$$\frac{\partial \xi^\mu}{\partial x^\nu} + L^\mu_{\nu\varrho}\, \xi^\varrho = -\alpha_i^{\cdot j}\, \lambda^i_\nu\, \lambda_j^\mu, \tag{3.6}$$

$$\xi^i_{,j} = -2\, \Omega^i_{jk}\, \xi^k - \alpha_j^{\cdot i}. \tag{3.7}$$

It is evident that if a space admits an infinitesimal motion ξ^μ it admits the finite group G_1 of motions generated by it; a direct verification of this assertion is to be found in § 7 below, where we obtain an explicit expression for an ennuple admitting a group G_1 of motions.

[9] Cf. H. Weyl, "Gruppentheorie und Quantenmechanik", 2nd ed., p. 107 (Leipsic 1930).

For a special infinitesimal motion we have $\alpha_i^{\cdot j} = 0$, and eq. (3.6) asserts that the field $\xi^\mu(x)$ defining such a special motion is a field of vectors which are "parallel" with respect to the transposed connection

$$\bar{L}^\mu_{\varrho\nu} = L^\mu_{\nu\varrho}. \tag{3.8}$$

R. A. P. S. The above results hold for a Riemannian absolute parallelism space provided the $a_i^{\cdot j}$ define a rotation (1.14) and the $\alpha_i^{\cdot j}$ $(= -\alpha_{\cdot i}^{j})$ an infinitesimal rotation (1.16). On eliminating the $\alpha_i^{\cdot j}$ from (3.5) with the aid of this anti-symmetry property we obtain the equations of Killing for the Riemannian metric (1.13):

$$\xi^\sigma \frac{\partial g_{\mu\nu}}{\partial x^\sigma} + g_{\sigma\nu}\frac{\partial \xi^\sigma}{\partial x^\mu} + g_{\mu\sigma}\frac{\partial \xi^\sigma}{\partial x^\nu} = 0. \tag{3.9}$$

The "special" motions, for which $a_i^{\cdot j}$ is the identity δ_i^j, can here more significantly be called *irrotational* motions.

4. **Conditions of integrability.** The conditions of integrability of the fundamental equations (3.5) are readily computed with the aid of (1.8), or they can be taken over directly from those obtained in § 2 for the equivalence problem by replacing $\bar{\lambda}_i^\mu(\bar{x})$, $a_i^{\cdot j}$ by $\lambda_i^\mu(x+\xi\cdot\delta\sigma)$, $\delta_i^j + \alpha_i^{\cdot j}\cdot\delta\sigma$ and expanding (cf. remarks concerning the conditions of integrability of (3.1) in § 3 above); they are

$$\Omega^i_{jk,m}\,\xi^m = -\alpha_m^{\cdot i}\,\Omega^m_{jk} + \alpha_j^{\cdot m}\,\Omega^i_{mk} + \alpha_k^{\cdot m}\,\Omega^i_{jm}, \tag{4.1}$$

$$\begin{cases} \Omega^i_{jk,m}\,\xi^m = -\alpha_m^{\cdot i}\,\Omega^m_{jk,l} + \cdots + \cdots + \alpha_l^{\cdot m}\,\Omega^i_{jk,m}, \\ \quad\vdots \qquad\qquad\qquad\qquad \vdots \\ \Omega^i_{jk,l_1\cdots l_\alpha m}\,\xi^m = -\alpha_m^{\cdot i}\,\Omega^m_{jk,l_1\cdots l_\alpha} + \cdots + \alpha_{l_\alpha}^{\cdot m}\,\Omega^i_{jk,l_1\cdots m}. \\ \quad\vdots \qquad\qquad\qquad\qquad \vdots \end{cases} \tag{4.2}$$

The necessary and sufficient conditions that the space T_n defined by λ_i^μ admit a group G_1 of motions is that there exist a constant matrix $\alpha_i^{\cdot j}$ and a non-negative integer N such that (1) *the first N sets of equations* (4.1), (4.2) *are compatible for the ξ^m as functions of the independent variables x and* (2) *all such solutions for these values of $\alpha_i^{\cdot j}$ satisfy the $(N+1)$st set.*

R. A. P. S. We need in this case only require that in addition the $\alpha_i^{\cdot j}$ satisfy the conditions (1.16) for an infinitesimal rotation.

5. **Commutator conditions.** Equations (3.5) are linear and homogeneous in the ξ^μ, their first derivatives and the constants $\alpha_i^{\cdot j}$; hence if the space admits the two infinitesimal motions

$$\delta A\colon\ \xi^\mu,\ \alpha_i^{\cdot j};\qquad \delta B\colon\ \eta^\mu,\ \beta_i^{\cdot j} \tag{5.1}$$

it also admits their sum

$$\delta A + \delta B: \quad \xi^\mu + \eta^\mu, \ \alpha_i^{\cdot j} + \beta_i^{\cdot j}. \tag{5.2}$$

But since these motions (5.1) are infinitesimal elements of a group the space must also admit their commutator $[\delta B, \delta A] \equiv \delta B \cdot \delta A - \delta A \cdot \delta B$ whose generator is

$$\zeta^\mu = [\xi\,\eta]^\mu \equiv \xi^\nu \frac{\partial \eta^\mu}{\partial x^\nu} - \eta^\nu \frac{\partial \xi^\mu}{\partial x^\nu}; \tag{5.3}$$

the fact that the rotations $a_i^{\cdot j}(\sigma)$ constitute a representation of the transformation group $G(x)$ tells us that the infinitesimal affine transformation of $G(\lambda)$ associated with this commutator is the commutator

$$\gamma_i^{\cdot j} = \beta_i^{\cdot k}\, \alpha_k^{\cdot j} - \alpha_i^{\cdot k}\, \beta_k^{\cdot j} \tag{5.4}$$

of the infinitesimal transformations $\alpha_i^{\cdot j}$, $\beta_i^{\cdot j}$. That the commutator (5.3), (5.4) does in fact define an infinitesimal motion of the space into itself is immediately shown by reducing the Jacobi identity

$$[[\xi\,\eta]\,\lambda_i] + [[\eta\,\lambda_i]\,\xi] + [[\lambda_i\,\xi]\,\eta] = 0 \tag{5.5}$$

with the aid of the fundamental equations (3.5) satisfied by ξ^μ, η^μ.

These considerations, which involve the fact that $G(\lambda)$ is a representation of $G(x)$, will be found an important tool in determining the spaces admitting automorphisms for which $G(x)$ is a given group. If the totality of representations are known for the group $G(x)$ in question the solutions of (5.4) are known explicitly—although not every such solution need lead to a solution of the geometrical problem (cf. § 11).

R. A. P. S. For motions in such a space it is again merely necessary to require that α_{ij} be anti-symmetric.

6. **The complete group admitted by a given space.** We now apply the results of §§ 4–5 to the problem of determining the complete group of motions admitted by a space of given structure, i. e. that group G_r of motions which contains all the automorphisms which are generated by the reiteration of infinitesimal elements. The order r can at most equal $n(n+1)$, for equations (3.6) define the first derivatives of ξ^μ linearly and homogeneously in terms of the n components ξ^μ themselves and the n^2 constants $\alpha_i^{\cdot j}$, whence the general solution can contain at most $n(n+1)$ arbitrary constants (e. g. the values of the ξ^μ at $x^\mu = x_0^\mu$ and the constants $\alpha_i^{\cdot j}$); a group of this greatest order will be said to be *maximal*. We denote by G_p that invariant sub-group of G_r consisting of all special motions; its order p is by the above at most equal to n—in which case it will be said to be a *maximal*

sub-group of special motions—and for $p = 0$ the only special automorphism with which we are here concerned is the identity.

We first discuss the case $0 < p \leqq n$ in which there exists a proper invariant sub-group G_p of special motions. This case arises if and only if the equations obtained by equating the left-hand sides of (4.1), (4.2) to zero are of m-rank $n - p$, for there then exist exactly p independent infinitesimal motions ξ^μ satisfying the fundamental equations (3.5) for $\alpha_i^{\cdot j} = 0$; each of these motions is of order zero at any point in space, as otherwise its components ξ^μ would vanish identically in virtue of the fundamental equations, and furthermore there exists no non-trivial homogeneous linear relation between the p sets with coefficients which may be functions of the x^μ. Hence the minimal invariant varieties of G_p are p-dimensional; in view of the interpretation of special motions given in § 3 above these minimal invariant varieties may be considered as generated by parallel propagation with respect to the transposed connection $\bar{L}^\mu_{\nu\varrho}$ (3.8) of the vectors ξ^μ in the $p-1$ parameter family of directions defined by the ξ^μ themselves. Now for each non-vanishing set of values of the $\alpha_i^{\cdot j}$ for which the solutions of the first α sets of (4.1), (4.2) satisfy the $(\alpha+1)^{\text{st}}$ set there exists exactly *one* additional (non-special) independent infinitesimal motion of G_r—for the difference of any two solutions for the same constants $\alpha_i^{\cdot j}$ is a special motion. Hence the number $q \leqq n^2$ of linearly independent sets of constants $\alpha_i^{\cdot j}$ which make equations (4.1), (4.2) compatible in the sense of § 4 determines the number of additional motions in the complete group G_r, and $r = p + q$.

The case in which there exists no special sub-group G_p arises when the m-rank of equations (4.1), (4.2) for $\alpha_i^{\cdot j} = 0$ is $\geqq n$. The order $q = r$ of the complete group G_r is then equal to the number q of linearly independent sets of constants $\alpha_i^{\cdot j}$ which make (4.1), (4.2) compatible, for each such set leads to a solution ξ^μ of (3.5), and if there existed two distinct solutions for the same $\alpha_i^{\cdot j}$ their difference would generate a special G_1 of motions.

R. A. P. S. Here the maximal group is of order $n(n+1)/2$, for the constants α_{ij} must now be anti-symmetric. The invariant sub-group G_p of special motions is the same G_p as in the general case above.

III. Determination of Spaces Admitting Certain Groups.

This final section is devoted to the determination of spaces admitting certain specified groups which are of interest either for the general theory of groups of motions in absolute parallelism spaces (§§ 7–9) or for any physical theory based on a Riemannian space-time admitting absolute parallelism (§§ 10–12).

7. **Group G_1. Reflection.** In order to determine a canonical form for a space admitting a group G_1 of motions we choose a coördinate system x^μ in which the generator of $G_1(x)$ assumes the canonical form $\xi^\mu = \delta^\mu_1$. Equation (3.5) then becomes

$$\frac{\partial \lambda_i^\mu}{\partial x^1} = \alpha_i^{\cdot j}\, \lambda_j^\mu \tag{7.1}$$

or, in matrix form,

$$\frac{\partial A^\mu}{\partial x^1} = A\, A^\mu \tag{7.2}$$

where A^μ is the matrix whose single column consists of the elements λ_i^μ $(i = 1, 2, \cdots, n)$ and A is the square matrix whose elements are $\alpha_i^{\cdot j}$. The solution of this matrix equation is[10]

$$A^\mu\,(x^1, x^2, \cdots, x^n) = e^{Ax^1} M^\mu\,(x^2, \cdots, x^n) \tag{7.3}$$

where the initial values $\mu_i^\mu\,(x^2, \cdots, x^n)$ $(= M^\mu$ for $i = 1, 2, \cdots, n)$ are arbitrary functions of their arguments and e^A is the matrix defined by

$$e^A = \mathbf{1} + A + \frac{1}{2!} A^2 + \cdots + \frac{1}{r!} A^r + \cdots. \tag{7.4}$$

For the direct verification, mentioned in § 3, of the assertion that this space admits the finite automorphisms of the group G_1 generated by ξ^μ we need merely note that (3.1) becomes

$$A^\mu\,(x^1 + \tau, x^2, \cdots, x^n) = A(\tau)\, A^\mu\,(x^1, x^2, \cdots, x^n)$$

for the finite transformation

$$x^1 \to x^1 + \tau, \qquad x^\alpha \to x^\alpha \quad (\alpha \neq 1) \tag{7.5}$$

of the group G_1, where $A(\tau)$ is the matrix of the $a_i^{\cdot j}(\tau)$. This equation is then satisfied in virtue of (7.3), where

$$A(\tau) = e^{A\tau} \tag{7.6}$$

is the finite affine transformation generated by the infinitesimal transformation A. It is to be noted that any quantity obtained from the λ_i^μ and their derivatives and which is invariant under change of ennuple—i. e. which carries no free Latin index—is independent of x^1. The coefficients of affine connection are, for example,

$$L^\mu_{\nu 1} = -\mu_\nu^i\, \alpha_i^{\cdot j}\, \mu_j^\mu, \qquad L^\mu_{\nu\alpha} = -\mu_\nu^i \frac{\partial \mu_i^\mu}{\partial x^\alpha} \quad (\alpha \neq 1). \tag{7.7}$$

[10] For a proof of the existence of e^A and the solution (7.3) see J. v. Neumann „Über die analytischen Eigenschaften von Gruppen linearer Transformationen und ihrer Darstellungen", Math. Zeits. 30, pp. 5–42 (1924); Weyl „Gruppentheorie und Quantenmechanik", p. 25.

In order that the spaces $x^1 = \text{const.}$ be autoparallel it is necessary and sufficient that it be possible to choose the above ennuple in such a way that $\mu_a^1 = 0$.

A problem related to the above is that in which the reflection

$$\mathbf{1'}: \quad x^1 \to -x^1, \qquad x^\alpha \to x^\alpha \quad (\alpha = 2, 3, \cdots, n) \tag{7.8}$$

in the hyper-surface $x^1 = 0$ constitutes an automorphism. The fundamental equations (3.1) are now

$$\begin{cases} \lambda_i^1(-x^1, x^2, \cdots, x^n) = -a_i^{\cdot j}\lambda_j^1(x^1, x^2, \cdots, x^n), \\ \lambda_i^\alpha(-x^1, x^2, \cdots, x^n) = a_i^{\cdot j}\lambda_j^\alpha(x^1, x^2, \cdots, x^n). \end{cases} \tag{7.9}$$

On choosing the ennuple in such a way that $\lambda_i^\mu = \delta_i^\mu$ at P_0: $x^\mu = 0$ we find that

$$a_i^{\cdot 1} = -\delta_i^1, \qquad a_i^{\cdot a} = \delta_i^a \qquad (a = 2, 3, \cdots, n).$$

The necessary and sufficient conditions that the space admit the reflection (7.8) is therefore that the ennuple may be chosen in such a manner that

$$\lambda_1^1 \to \lambda_1^1, \qquad \lambda_a^1 \to -\lambda_a^1, \qquad \lambda_1^\alpha \to -\lambda_1^\alpha, \qquad \lambda_a^\alpha \to \lambda_a^\alpha \tag{7.10}$$

on replacing x^1 by $-x^1$. The hyper-surface $x^1 = 0$ is autoparallel, as on it $\lambda_a^1 = 0$.

We now investigate the restrictions placed on the ennuple by requiring that it admit both the group (7.5) of translations and the reflection (7.8). No restrictions were placed on the orientation of the ennuple in deriving (7.3), so we are free to choose it as in the derivation of (7.10). This latter then requires that μ_a^1, μ_1^α vanish, and since at the point P_0 introduced above $\mu_i^\mu = \delta_i^\mu$ we find from (7.3) that

$$\lambda_i^\mu(x^1, 0, \cdots, 0) = \delta_i^\mu + \alpha_i^{\cdot\mu} x^1 + \cdots. \tag{7.11}$$

Since λ_1^1, λ_a^α are to remain unaltered on changing the sign of x^1 we must conclude that the corresponding elements $\alpha_1^{\cdot 1}$, $\alpha_a^{\cdot b}$ of A all vanish. These necessary conditions are also sufficient, for on expanding (7.3) we find that

$$\begin{cases} \lambda_1^1 = \mu_1^1 \cosh \alpha x^1, & (\alpha^2 = \alpha_1^{\cdot a}\alpha_a^{\cdot 1}), \\ \lambda_a^1 = \dfrac{\alpha_a^{\cdot 1}}{\alpha}\mu_1^1 \sinh \alpha x^1, & \lambda_1^\alpha = \dfrac{\alpha_1^{\cdot a}}{\alpha}\mu_a^\alpha \sinh \alpha x^1, \\ \lambda_a^\alpha = \mu_a^\alpha + \dfrac{\alpha_a^{\cdot 1}\alpha_1^{\cdot b}}{\alpha^2}\mu_b^\alpha(\cosh \alpha x^1 - 1) \end{cases} \tag{7.12}$$

and this ennuple satisfies (7.10). The above form of the ennuple is for the case $\alpha_1^{\cdot a}\alpha_a^{\cdot 1}>0$; the corresponding canonical forms for the cases in which this sum is less than or equal to zero are readily obtained from the above by replacing α by $i\alpha$ or allowing $\alpha\to 0$, respectively. The surfaces $x^1 = \text{const.}$ are autoparallel.

R. A. P. S. It is only necessary to require that the α_{ij} be antisymmetric in the above in order to obtain the corresponding results for a Riemannian absolute parallelism space. The $A(\tau)$ defined by (7.6) is a finite rotation generated by the infinitesimal rotation $\mathcal{A}$; a direct verification of the fact that this $A(\tau)$ defines a rotation is readily obtained on introducing the unipotent matrix ε whose elements are $\varepsilon_1^{\cdot j} = g_{ij}$ (1.12). For equations (1.14), (1.16) may be written

$$\varepsilon A\,\varepsilon A^* = \mathbf{1}, \qquad \mathcal{A}^* + \varepsilon\mathcal{A}\,\varepsilon = 0 \tag{7.13}$$

where the asterisk indicates the transposed matrix, and since it follows from (7.4) and the above that

$$e^{-A^*} = e^{\varepsilon A\varepsilon} = \varepsilon\, e^{A}\,\varepsilon$$

the desired result is established. The coefficients

$$g^{\mu\nu} = \mathcal{A}^{\mu^*}\,\varepsilon\,\mathcal{A}^{\nu} = M^{\mu^*}\,\varepsilon\,M^{\nu} \tag{7.14}$$

are of course independent of x^1, as is any quantity with no free Latin index formed from the λ_i^{μ}, $g_{\mu\nu}$ and their derivatives.

A slight change in the argument leading to (7.10) is now necessary, for as noted in § 1 the requirement that $\lambda_i^{\mu} = \delta_i^{\mu}$ at P_0 here demands that the coördinates x^{μ} be chosen in such a way that the metric assume the canonical form (1.12) at P_0. However, an examination of the argument concerning the reflection (7.8) shows that this is possible in virtue of the fact that we must have $g_{1\alpha} = 0$ on $x^1 = 0$.

8. **Maximal sub-group of special motions. Motions in group space.** In case the invariant sub-group of special motions is maximal, i.e. contains n independent infinitesimal motions, the conditions of integrability

$$\Omega^i_{jk,\,m}\,\xi^m = 0,$$

obtained from (4.1) by setting $\alpha_i^{\cdot j} = 0$, must be satisfied identically in the n components ξ^m; hence we must have

$$\Omega^i_{jk,m} = 0 \quad \text{or} \quad \Omega^i_{jk} = \text{const.} \tag{8.1}$$

But this is exactly the condition that our n-space T_n be a group manifold of an n-parameter continuous group H_n (not to be confused with the group G_r of motions admitted by the space!) of transformations[11]

$$\bar{y}^A = f^A(y^B, x^\mu) \qquad \begin{pmatrix} A, B = 1, 2, \cdots, N \\ \mu = 1, 2, \cdots, n \end{pmatrix} \tag{8.2}$$

for which

$$\frac{\partial \bar{y}^A}{\partial x^\mu} = \eta_i^A(\bar{y})\, \lambda_\mu^i(x). \tag{8.3}$$

The constants of composition c_{ij}^k of H_n are expressible in terms of the λ_i^μ in (8.3) by

$$c_{ij}^k = 2\lambda_{[i}^\mu \lambda_{j]}^\nu \frac{\partial \lambda_\mu^k}{\partial x^\nu} = 2\Omega_{ij}^k \tag{8.4}$$

and satisfy the relations

$$c_{(ij)}^k = 0, \qquad c_{[ij}^l c_{k]l}^m = 0 \tag{8.5}$$

in virtue of (1.5), (8.1) and the identity (1.6).[12]

For this case the problem thus becomes that of determining all automorphisms which preserve the structure of the group H_n.[13] The necessary and sufficient condition for such an automorphism is now merely that there exist a non-singular matrix $a_i^{\cdot j}$ such that

$$a_{\bar{i}}^{\cdot i}\, c_{\bar{j}\bar{k}}^{\bar{i}} = a_{\bar{j}}^{\cdot j}\, a_{\bar{k}}^{\cdot k}\, c_{jk}^i; \tag{8.6}$$

the corresponding condition for the existence of a one-parameter continuous group of such automorphisms is that there exist a matrix $\alpha_i^{\cdot j}$ (not necessarily non-singular) satisfying

$$\alpha_m^{\cdot i}\, c_{jk}^m = \alpha_j^{\cdot m}\, c_{mk}^i + \alpha_k^{\cdot m}\, c_{jm}^i. \tag{8.7}$$

We first consider the sub-group G_n of special motions. The fundamental equations (3.7) are, on setting $\alpha_i^{\cdot j} = 0$,

$$\xi_{,j}^i = -c_{jk}^i \xi^k; \tag{8.8}$$

but these are exactly the conditions that the vector ξ^μ define a (-)-constant field in the sense of Cartan and Schouten, and the n independent vectors

[11] For a systematic treatment of the geometry of group space see J. A. Schouten "Zur Geometrie der kontinuierlichen Transformationen", Math. Ann. 102, pp. 244–272 (1929); hereinafter referred to as "G. K. T." Cf. in particular pp. 249–250.

[12] G. K. T. p. 250, eqs. (12), (13); p. 256, eq. (48).

[13] Cf. E. Cartan and J. A. Schouten "On the Geometry of the Group Manifold of Simple and Semi-simple Groups", Amsterdam Akad. Wetensch. Proc. **29**, pp. 803–815 (1926); in particular p. 811. Hereinafter referred to as "G. G. M."

defining the (−)-equipollence are the generators of the second parameter group associated with H_n.[14] Our result is therefore: *an absolute parallelism space admitting a maximal invariant subgroup G_n of special motions is the group space of an n-parameter group H_n, and $G_n(x)$ is the second parameter group of H_n.*

The first parameter group associated with H_n, the generators of which are the λ_u^μ themselves, is also admitted as an n-parameter sub-group (in general non-special) of motions.[15] This result follows immediately from the fundamental equations (3.5) on writing (8.4) in the form

$$[\lambda_u \lambda_i]^\mu = c_{ui}^j \lambda_j^\mu; \tag{8.9}$$

$\lambda_u^{\cdot i}$ is therefore an infinitesimal motion for which

$$\alpha_i^{\cdot j} = c_{ui}^j. \tag{8.10}$$

The complete group G_r thus contains both the first and second parameter groups of H_n as transitive invariant sub-groups of order n; these two sub-groups possess no common sub-group if and only if the first parameter group contains no exceptional G_1, i. e. if and only if the u-rank of c_{ui}^j is n.

Cartan[16] has determined the automorphisms admitted by the group spaces of simple and semi-simple groups.

R. A. P. S. In this case the space is again the group space of an n-parameter group H_n and the (maximal) sub-group of irrotational motions is as in the general case the second parameter group of H_n. But the first parameter group of H_n no longer defines in general a sub-group of motions, as the constants α_{ij} must now be anti-symmetric; the condition that λ_u^μ define an infinitesimal motion is that

$$(\alpha_{ij} =) e_j c_{ui}^j = -e_i c_{uj}^i. \tag{8.11}$$

Hence the first parameter group defines an invariant sub-group of motions if and only if $c_{ijk} = e_i c_{jk}^i$ is completely anti-symmetric.

The complete group of isomorphisms in the group space of a simple or semi-simple group, under which the Riemannian metric (1.13) is invariant, is that sub-group of the isomorphisms discussed by Cartan for which the a_i^j define a rotation (1.14).

9. Maximal complete group. An absolute parallelism space which admits a maximal group $G_{n(n+1)}$ of motions must in particular admit a

[14] G. G. M., eq. (18), p. 807 (with $-c_{ij}^k$ in place of c_{ij}^k as pointed out by the authors on p. 946); G. K. T., eq. (64), p. 258 and eq. (27), p. 253.

[15] G. G. M., p. 812.

[16] E. Cartan, "Le principe de Dualité et la Théorie des Groupes Simples et Semi-simples", Bull. d. Sci. Math. 49, pp. 361–374. Eq. (8.6) is Cartan's fundamental eq. (6), p. 363.

maximal sub-group of special motions and must therefore be a group space, as shown above. But in addition the conditions of integrability (8.7) must be satisfied identically in the n^2 constants α_i^j; this requires that the constants c_{jk}^i satisfy the equations

$$\delta_m^i c_{jk}^l = \delta_j^l c_{mk}^i + \delta_k^l c_{jm}^i. \tag{9.1}$$

On contracting with respect to $i = n$ we find that

$$n c_{jk}^l = \delta_j^l c_{mk}^m + \delta_k^l c_{jm}^m$$

and on contracting again with respect to $l = j$ we find that $c_{km}^m = -c_{mk}^m = 0$, whence it follows from the above equations that

$$c_{jk}^i = 0. \tag{9.2}$$

But our space T_n is then the group space of an Abelian group H_n and is therefore flat—a coördinate system x^μ may be introduced in which the ennuple λ_i^μ may be taken as δ_i^μ, as is immediately evident from the equivalence conditions (2.3). The sub-group of special motions is then the group G_n of translations $x^\mu \to x^\mu + a^\mu$ and the complete group consists of these together with the full affine group G_{n^2}, as is apparent from the fundamental equations (3.1).

R. A. P. S. In order that the group admitted by a Riemannian absolute parallelism (group) space be of maximal order $n(n+1)/2$ the constants of composition $c_{jk}^i = e_i c_{ijk}$ must be such that (8.7) is satisfied identically for all anti-symmetric constants $\alpha_{ij} = e_j \alpha_i^j$. Condition (9.1) must then be replaced by

$$e_m (\delta_{li} c_{mjk} + \delta_{lj} c_{imk} + \delta_{lk} c_{ijm}) = e_l (\delta_{mi} c_{ljk} + \delta_{mj} c_{ilk} + \delta_{mk} c_{ijl}). \tag{9.3}$$

On contracting with respect to $m = j$ we find

$$c_{lik} + (n-2) c_{ilk} = e_l \delta_{li} c_{mk}^m \tag{9.4}$$

and on multiplying this equation by $e_l \delta_{li}$ we find that c_{mk}^m vanishes. Equation (9.2), together with the one obtained from it by interchanging i and l, then yields

$$(n-1)(n-3) c_{ilk} = 0. \tag{9.5}$$

The case $n = 1$ being of no interest, we see that unless $n = 3$ the constants c_{ijk} all vanish; the space is therefore as in the above the group space of an Abelian group, and the complete group $G_{n(n+1)/2}$ consists of the group G_n of all translations as the invariant sub-group of rotations and the group $G_{n(n-1)/2}$ of all rotations (1.14).

For the case $n = 3$ (9.4) requires that c_{ijk} be completely anti-symmetric; we may therefore write

$$c_{ijk} = \varepsilon_{ijk}/R \tag{9.6}$$

where $\varepsilon_{ijk} = \pm 1$ according as (ijk) is an even or an odd permutation of (123) and 0 otherwise, and R is an arbitrary constant. In order to obtain a canonical form for this case we first remark that the associated Riemannian space (1.13) must be of constant curvature $\varkappa$, since it admits the maximal group G_6 of motions, and may therefore be taken in the form[17]

$$ds^2 = \frac{e_i\, dx^i\, dx^i}{[1 + \varkappa e_i\, x^i\, x^i/4]^2}. \tag{9.7}$$

The six independent infinitesimal motions admitted by this space are

$$\begin{aligned} \xi^\mu_u &= \varepsilon_{\mu ui}\, e_i\, x^i && (u = 1, 2, 3), \\ \eta^\mu_u &= \delta^\mu_u \left(1 - \frac{\varkappa}{4} e_i\, x^i\, x^i\right) + \frac{\varkappa}{2} e_u\, x^\mu\, x^u \end{aligned} \tag{9.8}$$

and their commutators are

$$\begin{aligned} [\xi_u\, \xi_v]^\mu &= -\varepsilon_{uvw}\, e_w\, \xi^\mu_w, \qquad [\xi_u\, \eta_v]^\mu = -\varepsilon_{uvw}\, e_v\, \eta^\mu_w, \\ [\eta_u\, \eta_v]^\mu &= -\varepsilon_{uvw}\, \varkappa\, \xi^\mu_w \qquad (u, v, w = 1, 2, 3). \end{aligned} \tag{9.9}$$

Now it is readily shown that the G_6 generated by (9.8) possesses invariant sub-groups G_3 if and only if

$$\lambda^2 \equiv e_1 e_2 e_3 \varkappa \geqq 0 \tag{9.10}$$

and that for $\lambda^2 > 0$ there exist exactly two such invariant G_3 whose generators are

$$\zeta^\mu_u = \eta^\mu_u + \lambda e_u\, \xi^\mu_u, \qquad \zeta'^\mu_u = \eta^\mu_u - \lambda e_u\, \xi^\mu_u; \tag{9.11}$$

note that

$$\zeta^\mu_u\, e_u\, \zeta^\nu_u = \left(1 + \frac{\varkappa}{4} e_i\, x^i\, x^i\right)^2 e_\mu\, \delta^{\mu\nu} = g^{\mu\nu} \tag{9.12}$$

and similarly for ζ'^μ_u. Since c_{ijk} is completely anti-symmetric it follows from the remarks in § 8 that the λ^μ_u must themselves be the generators of an invariant sub-group G_3 of motions and may therefore be taken as linear combinations of the three vectors of either of the sets (9.11). But since by (9.12) each of these sets is itself an orthogonal ennuple in the Riemannian space defined by (9.7) we may take

$$\lambda^\mu_u = \zeta^\mu_u \quad \text{or} \quad \zeta'^\mu_u. \tag{9.13}$$

[17] R. G. pp. 239, 85.

It then follows from (8.9), (9.6) and (9.9) that

$$c_{uvw} = \frac{\varepsilon_{uvw}}{R} \quad \text{where} \quad \frac{1}{R} = \mp 2\lambda e_1 e_2 e_3 \tag{9.14}$$

and that the $\zeta_u'^{\mu}$ are the generators of the irrotational sub-group.

In summary, we have shown that the only non-flat Riemannian absolute parallelism spaces which admit a maximal group of motions are those 3-dimensional ones whose curvature $\varkappa$ satisfies the condition $e_1 e_2 e_3 \varkappa > 0$; the ennuple defining such a space may be taken in the canonical form (9.13), and it admits the G_6 of motions defined by (9.11). The two absolute parallelisms thus defined in these Riemannian spaces become the parallelisms of Clifford for a positive definite Riemannian space of positive constant curvature.[18] It follows immediately from eqs. (3.1) that these spaces do not admit the improper rotations unless $\varkappa = 0$.

10. **Applications. Stationary space-time.** The remaining sections are devoted to the discussion of certain Riemannian absolute parallelism spaces of interested to the unified field theory mentioned in the Introduction. For such spaces $n = 4$ and the signature is -2; on choosing x^α $(= 1, 2, 3)$ as the space-like coördinates and calling the time-like coördinate $x^0 = t$ instead of x^4 we may take

$$e_0 = 1, \qquad e_1 = e_2 = e_3 = -1. \tag{10.1}$$

Henceforth the Greek letters μ, ν, etc. and the Latin letters i, j, etc. run through the entire range 0, 1, 2, 3, whereas α, β, etc. and a, b, etc. are restricted to the values 1, 2, 3.

We shall understand by a *stationary space-time* one which admits a G_1 of motions whose infinitesimal generator ξ^μ is time-like. On introducing coördinates in which ξ^μ assumes the canonical form δ_0^μ we have as the canonical form for such a stationary space-time (cf. 7.3)

$$A^\mu(t, x) = e^{At} M^\mu(x) \quad \text{where} \quad A^* \varepsilon + \varepsilon A = 0. \tag{10.2}$$

The spaces $t =$ const. are autoparallel if and only if the coördinate system and the orientation of the ennuple may be further specialized in such a way that

$$\mu_a^0 = 0. \tag{10.3}$$

[18] Cartan and Schouten have also been led to the association of a Riemannian space with the group manifold of certain groups; see G. G. M. p. 811. See also remarks by E. T. Whittaker in "Parallelism and Teleparallelism in the Newer Theories of Space", London Math. Soc. Journ. 5, pp. 68–80 (1930); E. Bortolotti "On Parallelisms and Teleparallelisms in Curved Space", id. pp. 242–248. Further references to the literature are to be found in these papers.

If the space-time is also *reversible in time*, i. e. if it also admits the automorphism $t \to -t$, $x^\alpha \to x^\alpha$, then it must be of the form (cf. 7.12)

$$\lambda_0^0 = \mu_0^0 \cosh \alpha t, \qquad \lambda_a^0 = \frac{\alpha_a}{\alpha} \mu_0^0 \sinh \alpha t,$$

$$\lambda_0^\alpha = \frac{\alpha_a}{\alpha} \mu_a^\alpha \sinh \alpha t, \qquad \left(\alpha^2 = \sum_{a=1}^{3} \alpha_a^2\right), \tag{10.4}$$

$$\lambda_a^\alpha = \mu_a^\alpha + \frac{\alpha_a \alpha_b}{\alpha} \mu_b^\alpha (\cosh \alpha t - 1)$$

where the α_a are arbitrary real constants and the initial values μ of the λ_0^0, λ_a^α are arbitrary functions of the x^α. The spaces $t = \text{const.}$ are autoparallel.

If the motion is irrotational $A = 0$ and the ennuple

$$\lambda_i^\mu = \mu_i^\mu (x) \tag{10.5}$$

is independent of the time; we shall say that such a space is *static*. Previous writers[19] have employed this more restrictive assumption in dealing with space-times whose structure is unchanged in time. It is to be noted, however, that such a static universe is reversible if and only if

$$\mu_a^0 = 0, \qquad \mu_0^\alpha = 0; \tag{10.6}$$

this result may also be obtained from (10.4) on allowing $\alpha \to 0$. We shall in the following consider that the requirement of unvarying structure is met by the stationary fields (10.2), reserving for more restricted situations the stationary autoparallel fields (10.3), the reversible stationary fields (10.4), and the static fields obtained from them by the requirement that the motion be irrotational. As noted in § 7 above, any pure tensor (i. e. with no free Latin index) is independent of the time if the ennuple is defined by (10.2), so this requirement is sufficient for any unified field theory based on the λ_i^μ, g_{ij} alone in which physical quantities, such as the electro-magnetic field strengths, are described by such pure tensors.

11. **Spherical-symmetric space-time.** We now apply the results of our investigations to the *a priori* determination of the form of space-time suitable for describing the field of a spherically symmetric body — i. e. to the problem already discussed by Einstein and Mayer. Our assumptions, which are weaker than those of these authors, are 1) space-time is stationary and 2) spatially spherical-symmetric, i. e. it admits the full group G_3 of rotations about the center, the paths of which

[19] Einstein and Mayer, l. c. eq. (16), p. 112; McVittie, l. c. p. 142.

lie completely in the 3-spaces $t = \text{const.}$[20] In spite of our weaker assumptions, our results will be found to agree with those obtained by the former authors.

We first consider the restrictions imposed on the absolute parallelism space by assumption 2) alone. Let the paths of the time-like motion 1) define the t-axis; we then set up spatial "polar" coördinates consisting of r, some function of the normal geodesic distance from a fixed sphere r_0, and ϑ and φ, the co-latitude and longitude of the point in which the geodesic intersects the fixed sphere r_0. The group G_3 now consists merely of the group of rotations of the fixed sphere on itself; we find it convenient, however, to express the problem analytically in terms of the "Cartesian" coördinates x^α ($\alpha = 1, 2, 3$) which are defined in terms of r, ϑ and φ as usual in Euclidean 3-space. The generators of G_3 are, in terms of these coördinates x^α, the familiar rotations [cf. (9.8)]

$$\xi_u^0 = 0, \quad \xi_u^\alpha = \varepsilon_{\alpha u \beta} x^\beta \qquad (u, \alpha, \beta = 1, 2, 3). \tag{11.1}$$

We must now determine the associated Lorentz representation $A(\sigma)$ of this rotation group. The general theory of such representations tells us that there are but two single-valued 4-dimensional representations of the complete groups of proper 3-dimensional rotations: a) one which is equivalent to the identity taken four times and b) one which is equivalent to the identity taken once and the 3-dimensional rotation group itself.[21] In the latter case we may take as the representation $A(\sigma)$ the group G_3 itself,[22] which is a sub-group of the Lorentz group. The anti-symmetric matrix $A_u \varepsilon$ defining the infinitesimal transformation associated with (11.1) may therefore be taken in the form

$$\alpha_{ua0} = 0, \qquad \alpha_{uab} = M \varepsilon_{uab} \tag{11.2}$$

where $M = 0$, 1, corresponding to the two possibilities a), b).[23]

[20] Cf. treatment of the corresponding problem for the associated Riemannian space by J. Eiesland "The Group of Motions of an Einstein Space", Am. Math. Soc. Trans. 27, pp. 213–245 (1925).

[21] See, for example, Weyl, l. c.; in particular p. 128.

[22] For a formal proof of this assertion it is only necessary to show that the transformation matrix M in $A(\sigma) = M R(\sigma) M^{-1}$ can be taken as a Lorentz transformation (7.13), where $R(\sigma)$ is the 4-dimensional matrix describing the rotation sub-group G_3 of the Lorentz group in reduced form, for on subjecting the ennuple to the transformation M^{-1} we may take the new $A(\sigma) = R(\sigma)$. But it is readily shown that 1) $\varepsilon M^* \varepsilon M$ commutes with $R(\sigma)$ and must therefore be of the form $a\,1 + b\varepsilon$ where $a > |b|$ and 2) M may be replaced by a matrix $M(c\,1 + d\varepsilon)$, which does not affect $A(\sigma)$, in which the constants c, d may be so chosen that the new M satisfies the condition (7.13) $\varepsilon M^* \varepsilon M = 1$ for a Lorentz transformation—specifically $c = r\cos\theta$, $d = r\sin\theta$ where $\sin 2\theta = b/a$, $r^2 = a$.

[23] These results are of course obtainable directly from the commutation rules satisfied by the a_i^j, but not without considerable manipulation. An alternative method which gives

The fundamental equations (3.5) are now

$$\varepsilon_{\beta u\gamma}\, x^\gamma \frac{\partial \lambda_0^0}{\partial x^\beta} = 0, \tag{11.3}$$

$$\varepsilon_{\beta u\gamma}\, x^\gamma \frac{\partial \lambda_0^\alpha}{\partial x^\beta} - \lambda_0^\beta\, \varepsilon_{\alpha u\beta} = 0, \tag{11.4}$$

$$\varepsilon_{\beta u\gamma}\, x^\gamma \frac{\partial \lambda_a^0}{\partial x^\beta} = -M\, \varepsilon_{uab}\, \lambda_b^0, \tag{11.5}$$

$$\varepsilon_{\beta u\gamma}\, x^\gamma \frac{\partial \lambda_a^\alpha}{\partial x^\beta} - \lambda_a^\beta\, \varepsilon_{\alpha u\beta} = -M \varepsilon_{uab}\, \lambda_b^\alpha. \tag{11.6}$$

From (11.3) we conclude that

$$\lambda_0^0 = F(r, t), \qquad \text{where } r^2 = x^a x^a. \tag{11.7}$$

On multiplying (11.4) by x^u and summing we find that λ_0^α is of the form $E(x^\beta, t)\, x^\alpha$, and on substituting this result back in (11.4) that

$$\lambda_0^\alpha = E(r, t)\, x^\alpha. \tag{11.8}$$

Consider now the possibility a), i. e. the case $M = 0$. The remaining two sets of equations are, for given a, then the same as the first two and their solutions are therefore

$$\lambda_a^0 = D_a(r, t), \qquad \lambda_a^\alpha = C_a(r, t)\, x^\alpha;$$

however this is impossible, as in this case $|\lambda_i^\mu| = 0$ and the vectors cannot constitute an ennuple. Hence we must have b), $M = 1$; (11.5) is then essentially the same set of equations as (11.4), and its solution is

$$\lambda_a^0 = D(r, t)\, x^a. \tag{11.9}$$

Finally, on multiplying the last equation by x^u and summing we find

$$x^u (\varepsilon_{u\alpha\beta}\, \lambda_a^\beta + \varepsilon_{uab}\, \lambda_b^\alpha) = 0;$$

---

them immediately, and which is more allied to that employed by Einstein and Mayer, consists in choosing r in such a way that it is zero at the "center of symmetry", at which point we take $\lambda_i^\mu = \delta_i^\mu$; (11.2), for $M = 1$, then follows immediately from the fundamental equations (3.5) for (11.1) on setting $x^\alpha = 0$. The objection that the "point" $r = 0$ may be of the nature of the singularity in the Schwarzschild solution can be circumvented by applying this reasoning to a case in which $r = 0$ is a regular point and concluding that the result (11.2) points to the existence of a Lorentz transformation which establishes this result regardless of the nature of the "point" $r = 0$. We have preferred, however, to call on the theory of representations for this case in order to indicate the general method of attack for groups whose representations are known, but the simpler alternative method sketched in this foot-note will often be found useful—e. g. in the problem discussed by McVittie, l. c.

the solution of these algebraic equations is readily found to be

$$\lambda_a^\alpha = Ax^\alpha x^a + B\delta_a^\alpha + C\varepsilon_{\alpha a\beta}x^\beta, \tag{11.10}$$

and on substituting this result in (11.6) we find that A, B and C must be functions of r and t alone.

We have yet to discuss the result of imposing condition 1), which should determine the dependence of these six functions $A, \cdots, F$ on the time. Substituting (11.7)-(11.10) in the fundamental equations (3.5) for the transformation $t \to t+\tau$, $x^\alpha \to x^\alpha$ we find, on equating coefficients of the various powers of x^α, that in order that the space be stationary we must have

$$a_i^j(\tau) = \delta_i^j; \quad A = A(r), \cdots, F = F(r); \tag{11.11}$$

our result thus agrees completely with that of Einstein and Mayer.[24]

The most general space-time transformation which leaves the analytic expression of the complete group G_4 unaltered can be obtained from the four simple transformations

$$\text{a) } t \to at, \quad \text{b) } t \to t+\psi(r), \quad \text{c) } r \to \varphi(r), \quad \text{d) } x^\alpha \to r_\beta^\alpha x^\beta \tag{11.12}$$

where the r_β^α define a proper rotation. As shown by Einstein and Mayer[25] D and A can be made to vanish by transformations of the type (11.12b) and (11.12c), respectively—hence the coördinates can be so chosen that the spaces $t = \text{const.}$ are autoparallel. But in order that the space-time be reversible we must have $D = E = 0$, and this constitutes a real restriction. Finally, as noted by the previous authors, improper rotations are admitted if and only if $C = 0$—which can be seen immediately from (7.10) on noting that the space must then admit the reflection $x^1 \to -x^1$.

The associated Riemannian space is defined by the metric tensor $g_{\mu\nu}$ whose contravariant components are

$$\begin{aligned} g^{00} &= F^2 - D^2r^2, \quad g^{0\alpha} = [EF - D(Ar^2+B)]x^\alpha, \\ g^{\alpha\beta} &= -(B^2+C^2r^2)\delta^{\alpha\beta} - (A^2r^2 + 2AB - C^2 - E^2)x^\alpha x^\beta. \end{aligned} \tag{11.13}$$

12. **Space-times for relativistic cosmology.** As a final application of the theory developed in the previous sections we consider the Riemannian

[24] Einstein and Mayer, l. c. p. 114. To obtain the result $A(\tau) = 1$ in terms of the theory of representations it is only necessary to remark that $A(\tau)$ commutes with $R(\sigma)$ and must therefore be of the form $a1 + b\varepsilon$. But since it is also to be a Lorentz transformation $\varepsilon A^* \varepsilon A = 1$ or $ab = 0$, $a^2 + b^2 = 1$. Hence $A(\tau) = \pm 1$ or $\pm\varepsilon$, and since it is to be the identity for $\tau = 0$ we obtain the desired result.

[25] l. c. p. 114.

absolute parallelism spaces possessing the *a priori* symmetry properties required by the theory of relativistic cosmology. These requirements, which were set up by the author for Riemannian space,[26] are expressed by the assumption that space-time admits an intransitive group G_6 of spatial motions. This previous work shows that such a Riemannian space must be of the form

$$ds^2 = dt^2 - R^2(t)\, du^2 \tag{12.1}$$

where du^2 is the linear element of a positive definite 3-space of constant curvature, say $\varkappa$; the associated Riemannian space (1.13) may therefore be taken in the form (12.1). On choosing spatial coördinates x^α as in § 9 (for $e_u = +1$) we have

$$du^2 = \frac{dx^\alpha\, dx^\alpha}{[1 + \varkappa r^2/4]^2} \tag{12.2}$$

where $r^2 = x^\alpha x^\alpha$. The group G_6 admitted by this 3-space, defined by the generators ξ^μ_u, η^μ_u in (9.8) for $e_u = +1$, may be taken as the group admitted by the 4-space (12.1) provided we add $\xi^0_u = \eta^0_u = 0$ to the defining equations; the commutator rules satisfied by these 4-dimensional generators are then given by (9.9) on allowing μ to run through the range 0, 1, 2, 3.

Since our absolute parallelism space admits the rotational sub-group G_3 defined by (11.1) the results of § 11 apply to the present problem; in particular, the constants $\alpha^{\cdot j}_{ui}$ associated with the infinitesimal motions of this G_3 may be taken as in (11.2) and the λ^μ_i must be of the form (11.7)–(11.10). The commutation rules (9.9) for ξ^μ_u, η^μ_v lead, in accordance with (5.4) and (11.2), to the equations

$$\begin{aligned} \beta_{v0b}\, \varepsilon_{uba} &= \varepsilon_{uvw}\, \beta_{w0a}, \\ \beta_{vac}\, \varepsilon_{ucb} - \varepsilon_{uac}\, \beta_{vcb} &= \varepsilon_{uvw}\, \beta_{wab}, \end{aligned} \tag{12.3}$$

for the constants $\beta^{\cdot j}_{ui}$ associated with the motions η^μ_u. The solution of these equations is readily found to be

$$\beta_{ua0} = M\delta_{ua}, \qquad \beta_{uab} = N\varepsilon_{uab} \tag{12.4}$$

and the additional equations for $\beta^{\cdot j}_{ui}$ arising from the commutation rules for η^μ_u, η^μ_v in (9.9) are satisfied by (11.2), (12.4) provided

$$MN = 0, \qquad -M^2 + N^2 = \varkappa. \tag{12.5}$$

We consider first the case in which $M = 0$, $N^2 = \varkappa \geqq 0$. The fundamental equations (3.5) for the infinitesimal motion η^μ_u then yield, on

[26] H. P. Robertson "On the Foundations of Relativistic Cosmology", Nat. Acad. Sci. Proc. 15, pp. 822–829; in particular p. 824.

substituting the values of λ_i^μ given by (11.7)–(11.10), the following results. From (3.5) for $\mu = 0$, $i = 0$ we find that $F = F(t)$; for $\mu = \alpha$, $i = 0$ that $E = 0$; and from $\mu = 0$, $i = a$ that $D = 0$. On comparing (11.13) with (12.1) we now see that $F^2 = 1$, and we may therefore take $F = 1$. The remaining equations, for $\mu = \alpha$, $i = a$, now assert that the 3-space defined by λ_a^α admits the motions η_u^μ, and by (11.6) it also admits the motions ξ_u^μ. But since E, and therefore λ_0^α, vanish $R(t)\,\lambda_a^\alpha$ is an orthogonal ennuple for the 3-space (12.2) and the results of § 9 show that it may be taken as ζ_a^α or $\zeta_a'^\alpha$ [cf. (9.13)]—the possibility that $R(t)\,\lambda_a^\alpha$ is obtained from the ζ_a^α or $\zeta_a'^\alpha$ by an orthogonal transformation with coefficients which are functions of t being eliminated by comparison with (11.10). This possibility thus leads to the result

$$\lambda_0^0 = 1, \qquad \lambda_a^0 = 0, \qquad \lambda_0^\alpha = 0,$$

$$R(t)\,\lambda_a^\alpha = \delta_a^\alpha\left(1 - \frac{\varkappa}{4}r^2\right) + \frac{\varkappa}{2}x^\alpha x^a \pm \sqrt{\varkappa}\,\varepsilon_{\alpha a\beta}\,x^\beta \tag{12.6}$$

for $\varkappa \geqq 0$. This space admits the improper rotations if and only if $\varkappa = 0$.

The other possibility is that in which $N = 0$, $-M^2 = \varkappa \leqq 0$. Equations (3.5) for $\mu = 0$, $i = i$ now yield the independent equations

$$\frac{\partial F}{\partial r}\left(1 + \frac{\varkappa}{4}r^2\right) - MDr = 0,$$

$$D\left(1 - \frac{\varkappa}{4}r^2\right) - MF = 0$$

for the functions D and F. These equations, together with the condition

$$F^2 - D^2 r^2 = 1$$

obtained by comparing (12.1) and (11.13), allow us to write

$$F = \frac{1 - \varkappa r^2/4}{1 + \varkappa r^2/4}, \qquad D = \frac{M}{1 + \varkappa r^2/4}. \tag{12.7}$$

The fundamental equations (3.5) for $\mu = \alpha$, $i = 0$ become

$$C = 0, \qquad E\left(1 + \frac{\varkappa}{4}r^2\right) - MB = 0,$$

$$\frac{1}{r}\frac{\partial E}{\partial r}\left(1 + \frac{\varkappa}{4}r^2\right) - E\frac{\varkappa}{2} - MA = 0,$$

and these equations, together with the condition

$$B^2 + C^2 r^2 = \frac{1}{R^2(t)}\left(1 + \frac{\varkappa}{4}r^2\right)^2$$

obtained from (12.1) and (11.13), allow us to solve for the remaining functions

$$E = \frac{M}{R}, \qquad A = -\frac{\varkappa}{2R}, \qquad B = \frac{1}{R}\left(1+\frac{\varkappa}{4}r^2\right). \tag{12.8}$$

The ennuple defined by (12.7), (12.8) is now found to satisfy the remaining equations of the set (3.5) without further restriction. The absolute parallelism space defined by this ennuple admits the improper rotations, as is readily seen from the conditions (7.10) for the improper rotation $x^1 \to -x^1$.

Previous work has shown that the Riemannian space-time (12.1) is *stationary*, i. e. admits a time-like motion ξ^μ for which ξ^0 is a function of t alone, if and only if[27]

$$R(t) = R_0 e^{at}, \qquad \varkappa a = 0, \tag{12.9}$$

and that furthermore we may take $\xi^0 = 1$. The possibility $a = 0$ leads to cases in which $R = $ const. which are obviously stationary; (12.6) then corresponds to the original Einstein cosmology, but it is to be noted that the absolute parallelism space-time becomes flat if we require that it admit the improper rotations in addition. The second possibility, in which $\varkappa = 0$, $R = R_0 e^{at}$, gives rise to a stationary Riemannian absolute parallelism space admitting the motion

$$\xi^0 = 1, \qquad \xi^\alpha = -ax^\alpha; \qquad \alpha_i^{\cdot j} = 0 \tag{12.10}$$

and whose associated Riemannian space-time, of constant curvature $-1/a^2$, is that upon which the de Sitter cosmology is based.

[27] Robertson, l. c. p. 825. It is to be noted that the space-time coördinates here employed are chosen with reference to the physical content of the world, and that it would therefore be an essential restriction to choose the motion ξ^μ in the canonical form employed in § 10.

ON DUALITY AND INTERSECTION CHAINS IN COMBINATORIAL ANALYSIS SITUS.*

By J. H. C. Whitehead.

1. **Introductory.** This note is intended as a contribution to the formal exposition of combinatorial analysis situs. We do not attempt to cover much ground, but are content with a proof of the fundamental relation

$$F(C_p \cdot C_q^*) = C_p \cdot F(C_q^*) + (-1)^{n-q} F(C_p) \cdot C_q^*,$$

which is due to S. Lefschetz.[1] An account of the geometry underlying this note is to be found in his Colloquium Lectures.[2] The purely combinatorial theory has been developed in a series of articles by J. W. Alexander and by M. H. A. Newman,[3] further references to which are made in "Topology". We also call attention to a paper by Newman[4] and to Chap. II of a mémoire by G. de Rahm,[5] in which intersection chains are defined by means of looping coefficients without reference to the dual complex.

2. **Derived chains.** Let M_n be an orientable and oriented manifold

$$\frac{1}{(n+1)!} c^{\alpha_0 \cdots \alpha_n} x_{\alpha_0} \cdots x_{\alpha_n} \quad (\Sigma \text{ understood for } \alpha = 0, 1, \cdots N), \tag{2.1}$$

where the coefficients c are -1, 0 or $+1$, and are anti-symmetric in their indices. The symbols x obey the multiplication law $xx' = -x'x$. All relations will be relative, in the sense of Lefschetz, to $F(M_n)$, the boundary of M_n. Formally this is the same as assuming $F(M_n) = 0$. For non-orientable manifolds take everything mod. 2.

The first derived complex $D(M_n)$ is composed of cells

$$\pm x_{\alpha_0} x_{\alpha_0 \alpha_1} \cdots x_{\alpha_0 \cdots \alpha_n}, \tag{2.2}$$

where $x_{\alpha_0 \cdots \alpha_p}$ is the new vertex[6] associated with the p-cell $x_{\alpha_0} \cdots x_{\alpha_p}$. It follows by an easy induction that the cell (2.2), with the $+$ sign, is similarly oriented to the cell $x_{\alpha_0} \cdots x_{\alpha_n}$. Therefore

* Received December 8, 1931.

[1] Trans. American Math. Soc. 28 (1926), pp. 1–49.

[2] Lefschetz, Topology, New York, 1930. See, in particular, the beginning of Chapters III and IV.

[3] See for instance, Alexander, Annals of Math. 31 (1930), pp. 292–320; and Newman, Proc. Royal Academy of Amsterdam, 29 (1926), pp. 611 and 627.

[4] Proc. Cambridge Phil. Soc. 27 (1931), pp. 497–507.

[5] Journal de Math. p. et a. 10 (1931), pp. 115–200.

[6] This convenient notation was, I believe, introduced by Alexander.

$$D(M_n) = c^{\alpha_0 \cdots \alpha_n} x_{\alpha_0} x_{\alpha_0 \alpha_1} \cdots x_{\alpha_0 \cdots \alpha_n}, \tag{2.3}$$

where the c's are the same as those in (2.1).

Let

$$C_p = \frac{1}{(p+1)!} t^{\alpha_0 \cdots \alpha_p} x_{\alpha_0} \cdots x_{\alpha_p}$$

be any chain, the t's being positive, negative, or zero, and anti-symmetric in their indices. Its boundary is given by

$$F(C_p) = \frac{1}{p!} \Sigma_\beta t^{\beta \alpha_1 \cdots \alpha_p} x_{\alpha_1} \cdots x_{\alpha_p},$$

where β, like all Greek indices, runs from 0 to N. A useful formula follows from the fact that the boundary of $D(C_p)$ is the derived complex of $F(C_p)$. That is to say

$$FD(C_p) = DF(C_p) = \Sigma_\beta t^{\beta \alpha_1 \cdots \alpha_p} x_{\alpha_1} x_{\alpha_1 \alpha_2} \cdots x_{\alpha_1 \cdots \alpha_p}. \tag{2.4}$$

3. **Dual chains.** The complement of an r-cell in M_n,

$$E_r = x_{\alpha_0} \cdots x_{\alpha_r},$$

is given by

$$\frac{1}{(n-r)!} c^{\alpha_0 \cdots \alpha_r \lambda_1 \cdots \lambda_q} x_{\lambda_1} \cdots x_{\lambda_q}, \qquad (q = n-r) \tag{3.1}$$

and the chain dual to E_r by

$$E_q^* = \xi c^{\alpha_0 \cdots \alpha_r \lambda_1 \cdots \lambda_q} \xi_{\lambda_1} \xi_{\lambda_1 \lambda_2} \cdots \xi_{\lambda_1 \cdots \lambda_q}, \tag{3.2}$$

where

$$\begin{cases} \xi = x_{\alpha_0 \cdots \alpha_r}, \\ \xi_{\lambda_1 \cdots \lambda_j} = x_{\alpha_0 \cdots \alpha_r \lambda_1 \cdots \lambda_j}. \end{cases} \tag{3.3}$$

The first derived chain of (3.1) is given by

$$S_{q-1} = c^{\alpha_0 \cdots \alpha_r \lambda_1 \cdots \lambda_q} x_{\lambda_1} \cdots x_{\lambda_1 \cdots \lambda_q},$$

and by the definition of a manifold is a sphere, rel. $F(M_n)$ (whether a "true" or a "combinatory" sphere is irrelevant to this discussion). The congruence $x \to \xi$ carries xS_{q-1} into E_q^* and therefore the latter is a cell.

An orientation for any two of the chains M_n, E_r and E_q^* determines an orientation for the third. For if $\alpha_0 \cdots \alpha_r \lambda_1 \cdots \lambda_q$ are indices such that

$$c^{\alpha_0 \cdots \alpha_r \lambda_1 \cdots \lambda_q} \neq 0,$$

these three chains may be oriented by assigning a sense to the orders $\alpha_0 \cdots \alpha_r \lambda_1 \cdots \lambda_q$, $\alpha_0 \cdots \alpha_r$ and $\lambda_1 \cdots \lambda_q$ respectively. An order for $\lambda_1 \cdots \lambda_q$ determines an orientation for $F(E_q^*)$, and (3.2) fixes the position of the vertex[7] ξ.

4. **The intersection of chains.** If

$$E_p = x_{\beta_0} \cdots x_{\beta_p}$$

is any cell in M_n, then[8]

$$D(E_p) = \delta^{\lambda_0 \cdots \lambda_p}_{\beta_0 \cdots \beta_p} x_{\lambda_0} \cdots x_{\lambda_0 \cdots \lambda_p}. \tag{4.1}$$

For (4.1) and (3.2) to have a vertex in common it is necessary and sufficient that the α's are contained among the β's. That is to say $p \geqq r$ and E_r must be incident with E_p. For simplicity of notation let

$$E_r = x_0 x_1 \cdots x_r,$$
$$E_p = x_0 \cdots x_r x_{r+1} \cdots x_p.$$

Then the intersection $E_p \cdot E_q^*$ is given by

$$E_p \cdot E_q^* = \xi \theta^{\lambda_1 \cdots \lambda_s} \xi_{\lambda_1} \cdots \xi_{\lambda_1 \cdots \lambda_s}, \tag{4.2}$$

where $s = p - r$, the ξ's are given by (3.3), and

$$\theta^{\lambda_1 \cdots \lambda_s} = \delta^{\lambda_1 \cdots \lambda_s}_{r+1 \cdots p}.$$

It follows by arguments similar to those used in § 3 that $E_p \cdot E_q^*$ is a cell, and that an orientation for any three of M_n, E_p, E_q^* and $E_p \cdot E_q^*$ determines an orientation for the fourth. This is in accordance with the conditions stated in Chap. IV of "Topology". We suppose the cells in question to be positively oriented as written. From (4.2) and (2.4) we have

$$F(E_p \cdot E_q^*) = \theta^{\lambda_1 \cdots \lambda_s} \xi_{\lambda_1} \cdots \xi_{\lambda_1 \cdots \lambda_s} - \xi \Sigma_\sigma \theta^{\sigma\lambda_2 \cdots \lambda_s} \xi_{\lambda_2} \cdots \xi_{\lambda_2 \cdots \lambda_s}, \tag{4.3}$$

[7] This formalism should be convenient for a discussion of the Poincaré duality relations. For example, the relation

$$D(M_n) = D(M_n^*)$$

is evident when we set $r = 0$ in (3.2), replace ξ by x according to (3.3), and sum over α_0. We then have

$$D(M_n^*) = \Sigma_\alpha E_n^{*\alpha} = c^{\alpha_0 \cdots \alpha_n} x_{\alpha_0} \cdots x_{\alpha_0 \cdots \alpha_n},$$

which is identical with (2.3).

[8] The "generalized Kronecker δ" is 0 unless $\lambda_0 \cdots \lambda_p$ are obtained by permuting $\beta_0, \cdots, \beta_p$, and is $+1$ for even, -1 for odd permutations.

while from (3.2) we have

$$F(E_q^*) = c^{1\cdots r\lambda_1\cdots\lambda_q}\,\xi_{\lambda_1}\cdots\xi_{\lambda_1\cdots\lambda_q},$$

since $\partial E_q^*/\partial\xi$ is a sphere and its boundary vanishes. Therefore

$$E_p\cdot F(E_q^*) = \theta^{\lambda_1\cdots\lambda_s}\,\xi_{\lambda_1}\cdots\xi_{\lambda_1\cdots\lambda_s}. \tag{4.4}$$

We also have, from (2.4)

$$F(E_p) = \Sigma_\sigma\,\delta_{01\cdots p}^{\sigma\varrho_1\cdots\varrho_p}\,x_{\varrho_1}\cdots x_{\varrho_1\cdots\varrho_p},$$

and

$$\begin{aligned} F(E_p)\cdot E_q^* &= (-1)^{r+1}\,\xi\,\Sigma_\sigma\,\delta_{0\cdots r\cdots p}^{0\cdots r\sigma\lambda_2\cdots\lambda_s}\,\xi_{\lambda_2}\cdots\xi_{\lambda_2\cdots\lambda_s}, \\ &= (-1)^{r+1}\,\xi\,\Sigma_\sigma\,\theta^{\sigma\lambda_2\cdots\lambda_s}\,\xi_{\lambda_2}\cdots\xi_{\lambda_2\cdots\lambda_s}, \end{aligned} \tag{4.5}$$

the factor $(-1)^{r+1}$ being present because σ, as an index on δ, has been shifted $r+1$ places to the right. From (4.3), (4.4) and (4.5) we have

$$F(E_p,\ E_q^*) = E_p\cdot F(E_q^*)+(-1)^{n-q}\,F(E_p)\cdot E_q^*,$$

which leads to the fundamental relation for chains

$$F(C_p\cdot C_q^*) = C_p\cdot F(C_q^*)+(-1)^{n-q}\,F(C_p)\cdot C_q^*. \tag{4.6}$$

ON SEPARABLE SPACES.[1]

By S. Lefschetz.

We owe to E. H. Moore the introduction of a very general type of double sequence $\{P^{i\alpha}\}$ which he called a *development*. With such sequences there may be associated certain abstract spaces investigated at length by Chittenden and Pitcher with noteworthy results particularly in connection with the problem of metrization. They dealt at considerable length with so-called *regular* developments of a space $\mathfrak{R}$. In a regular development the P's are neighborhoods of $\mathfrak{R}$ and α has a finite range for every i. Our first object in the present paper is to investigate a type of development called *normal* whose sets are subjected to more stringent conditions of convergency than with the regular type. It turns out, however, that every separable metric space possesses normal developments. As a consequence they seem to be just what is needed for the treatment of many questions on separable metric.

Making use in part of normal developments, we have put on a solid basis the theory of the order for separable spaces, and in particular, completely extended to them the fundamental order theorem (Lebesgue order theorem). It was then a simple matter to prove that every separable metric space can be mapped topologically on a compact metric space of the same dimension. As a consequence certain mapping theorems that we have obtained previously for compact spaces[2] hold for separable spaces. By means of

[1] Received January 13, 1932. The results of the present paper were communicated in a Note to the Proceedings of the National Academy of Sciences, vol. 18, February 1932.

[2] See our paper *On Compact Spaces*, these Annals, vol. 32 (1931), pp. 521–538, also Nöbiling, Math. Ann., vol. 104 (1930), pp. 71–80, and a recent paper by Pontrjagin-Tolstova, Math. Ann., vol. 105 (1931), pp. 734–746. Regarding our paper, the sets G of No. 3 are better introduced as follows: Define first the sets

$$[F^{\alpha_0} \ldots\ldots F^{\alpha_k}] = F^{\alpha_0} \ldots\ldots F^{\alpha_k} - \Sigma F^{\beta}, \qquad \beta \text{ not an } \alpha_i, \tag{1}$$

then introduce closed sets G such that

$$G^{\alpha_0 \cdots \alpha_k} \supset [F^{\alpha_0} \ldots\ldots F^{\alpha_k}] - \Sigma G^{\alpha_0 \cdots \alpha_k \beta_1 \cdots \beta_h}, \qquad \beta_j \text{ not an } \alpha_i, \tag{2}$$

while $F^{\beta} \cdot G = 0$. We now select the G's so as to satisfy IV, VI, VII loc. cit., with E in VII replaced by [ ]. For $k = n$ the sets [ ] are closed and do not intersect, hence we can take for the G's closures of non-intersecting neighborhoods of the [ ]'s. Let everything go through down to $k+1$. On the one hand no two [ ] with $k+1$ terms meet; on the other hand $[F^{\alpha_0} \ldots\ldots F^{\alpha_k}]$ meets only sets $G^{\alpha_0 \cdots \alpha_k \beta_1 \cdots \beta_h}$. Hence we can find suitable G's

our results it is a simple matter to complete a noteworthy intersection theorem due to Menger and Hurewicz.[3]

§ 1. Generalities and Two Lemmas.

1. We shall denote our fundamental separable metric space by $\mathfrak{R}$, its points by $x, y, \cdots$, its distance-function by $d(x, y)$. If A, B are subsets of $\mathfrak{R}$, by $A \in B$ is meant that the closure, $\bar{A}$, of A is a subset of B. If $\sigma = \{A^\alpha\}$, $\sigma' = \{A'^\beta\}$ are two aggregates of subsets of $\mathfrak{R}$, we write $\sigma \subset \sigma'$ or $\sigma \in \sigma'$, whenever respectively every A or every $\bar{A}$ is a subset of an A'. We shall also denote by $\bar{\sigma}$ the aggregate $\{\bar{A}^\alpha\}$ consisting of the closures of the sets of σ.

An open subset of $\mathfrak{R}$ will be generically designated by the letters U, V, W with indices. An aggregate $\{U^i\}$, such that every open set of $\mathfrak{R}$ is a sum of sets U^i, is called a *base* of $\mathfrak{R}$. Separability is defined by the possession of an enumerable base.

A *covering* of $\mathfrak{R}$ is an aggregate of subsets $\{A\}$, whose sum is $\mathfrak{R}$; its *mesh* is the maximum diameter of the A's. The covering is *finite* or *infinite* according as the number of A's is finite or infinite. Of basic importance for the sequel are the finite coverings by open sets (= f. c. o. s.).

2. We recall that according to Urysohn $\mathfrak{R}$ can be mapped topologically on the Hilbert parallelotope $\mathfrak{H}$:

$$|y_i| < \frac{1}{i}, \qquad i = 1, 2, \cdots, +\infty,$$

which is itself compact and metric. The mapping is obtained as follows: Given any two closed subsets A, B of $\mathfrak{R}$, whose intersection $A \cdot B = 0$, there exists a continuous single-valued function $f(x)$ of the point x on $\mathfrak{R}$ such that

$$f(x) = 0 \text{ on } A, \quad f(x) = 1 \text{ on } B, \quad 0 < f(x) < 1 \text{ on } \mathfrak{R} - A - B.$$

We will call $f(x)$ a *characteristic function* of the pair A, B. Let $\{V^n\}$ be an enumerable base, and consider all the pairs (V^p, V^q) such that $V^q \in V^p$. We have $\bar{V}^q \cdot \mathfrak{R} - V^p = 0$, and since the two intersecting sets are closed they possess a characteristic function. The totality of these functions is enumerable; range them in a certain order and call them $f_i(x)$, $i = 1, 2, \cdots, +\infty$. The subset R of $\mathfrak{H}$ represented by

by taking closures of neighborhoods of the sets at the right in (2). The proof of the Lemma I is then carried out as before except that on p. 523 T and U should be interchanged.

[3] For accurate references to the literature the reader is asked to consult the recent article by Tietze-Vietoris, Encykl. der Math. Wiss. III A B 13, which appeared prior to the papers quoted in the preceding footnote. The notation and terminology of our paper are those now generally current in Topology.

$$y_i = \frac{f_i(x)}{i}, \qquad i = 1, 2, \cdots, +\infty, \tag{1}$$

is the required topological image of $\mathfrak{R}$.

3. LEMMA I. *Given an enumerable aggregate of non-intersecting pairs of closed sets* $\{(A^i, B^i)\}$, *the metric can be so chosen that every* $d(A^i, B^i) > 0$.

Let $g_i(x)$ be the characteristic function for the ith pair, and let the f's be as in (1). The equations

$$y_{2i} = \frac{g_i(x)}{2i}, \qquad y_{2i+1} = \frac{f_i(x)}{2i-1}, \qquad x \subset \mathfrak{R}, \tag{2}$$

represent a subset R of $\mathfrak{H}$. By paraphrasing Urysohn's reasoning in the theorem just recalled, it is shown that R is homeomorphic to $\mathfrak{R}$. Let us adopt for $\mathfrak{R}$ the metric of R, i. e. take as the distance of any two points of $\mathfrak{R}$ that of their images on R. If $x \subset A^i$, $x' \subset B^i$, we have $d(x, x') \geqq \left| \frac{g_i(x) - g_i(x')}{2i} \right| = \frac{1}{2i}$, which proves the Lemma.

It is hardly necessary to observe that when $\mathfrak{R}$ is compact the Lemma holds with the initial metric.

LEMMA II. *Given a f. c. o. s.* $\Sigma = \{U^\alpha\}$, *the metric can be so chosen that there exists an* η, *such that every subset of* $\mathfrak{R}$ *whose diameter* $< \eta$, *is a subset of some* U. *When* $\mathfrak{R}$ *is compact this holds without modifying the metric.*

We make use of the following property due to Menger.[4] *For every* U^α *there is a* $V^\alpha \in U^\alpha$ *such that* $\sigma = \{V^\alpha\}$ *is also a covering.* If we set $A^\alpha = \mathfrak{R} - U^\alpha$, a closed set, we have $\bar{V}^\alpha \cdot A^\alpha = 0$. By Lemma I we can choose a metric such that $d(\bar{V}^\alpha, A^\alpha) > 0$ for every α. These distances being in finite number they will have a lower bound $\eta > 0$. Let B be any set of diameter $< \eta$. Since σ is a covering B meets at least one V, say V^α. Since $d(V^\alpha, A^\alpha) \geqq \eta$, we will have $B \cdot A^\alpha = 0$, hence $B \subset U^\alpha$, which proves Lemma II for the separable case. When $\mathfrak{R}$ is compact we always have $d(\bar{V}^\alpha, A^\alpha) > 0$ without modifying the metric, and this completes the proof.

LEMMA III. *Given a finite covering by closed sets* $\sigma = \{A^\alpha\}$, *the metric can be so chosen that there exists an* $\eta > 0$ *such that if* $A^{\alpha_0} \cdots A^{\alpha_p} = 0$ *no set of diameter* $< \eta$ *meets all the sets* A^{α_i}. *When* $\mathfrak{R}$ *is compact this holds without modifying the metric.*

An equivalent property is: *The largest distance* $\delta(x)$ *from a point* x *to one of the sets* A^{α_i} *has a positive lower bound* η *as* x *ranges over* $\mathfrak{R}$.

Consider first a finite number of f. c. o. s. $\Sigma^i = \{U^{i\alpha}\}$, $i = 1, 2, \cdots, k$. I say that the metric can be so chosen that Lemma II holds simultaneously for all the Σ's. For the sets $U^{1\alpha} \cdot U^{2\beta} \cdots U^{k\gamma}$ constitute a f. c. o. s. Apply

[4] *Dimensionstheorie*, p. 160.

Lemma II to it and let η be as in the lemma. Then if diam $B<\eta$, there is a $U^{1\alpha}\cdot U^{2\beta}\dots U^{k\gamma}\supset B$, and hence a $U^{i\delta}\supset B$ for every i.

Let now $U^{\alpha}=\mathfrak{R}-A^{\alpha}$, an open set. We have

$$A^{\alpha_0}\dots A^{\alpha_p}=\Pi(\mathfrak{R}-U^{\alpha_i})=\mathfrak{R}-(U^{\alpha_0}+\dots+U^{\alpha_p}).$$

Hence if the sets A^{α_i} do not intersect, their complements U^{α_i} constitute a f. c. o. s. The number of the latter being finite, let the metric be so chosen that Lemma II applies simultaneously to all, with η as in the lemma. If diam $B<\eta$, there is then a $U^{\alpha_i}\supset B$ and hence $B\cdot A^{\alpha_i}=0$, which proves Lemma III. For the same reasons as previously, in the compact case no modification of metric is called for.

§ 2. Normal Developments.

5. By a *decreasing sequence of open sets* we shall understand a sequence of open sets $\{U^n\}$ such that $U^{n+1}\Subset U^n$ for every n. For such a sequence $\Pi U^n=\Pi\bar{U}^n$, and the intersection $\neq 0$ when $\mathfrak{R}$ is compact. Whether $\mathfrak{R}$ is compact or not, let the intersection exist and denote it by A. This set is necessarily closed since it is the intersection of an enumerable group of closed sets. If every open set $V\supset A$ contains all but a finite number of the sets U, we say that the sequence *converges* to A and write $\{U^n\}\to A$. When $A=0$ we say that the sequence converges to zero.

It is now a simple matter to define a *normal development* Δ. It consists of a sequence $\{\Sigma^i\}=\{\{U^{i\alpha}\}\}$ of f. c. o. s. with the following properties:

(a) $\{\{U^{i\alpha}\}\}$ is a base;

(b) every $\bar{U}^{i+1,\alpha}$ is on a $U^{i\beta}$;

(c) if $U^{i\alpha}\Subset U^{j\beta}$, then for i above a certain value, every $\bar{U}^{h\gamma}$ meeting $\bar{U}^{i\alpha}$ is likewise on $U^{j\beta}$;

(d) every decreasing sequence $\{U^{n_i\alpha_i}\}$ converges to zero or to a point.

Evidently (a) implies separability. It is important to observe that it may be replaced by:

(a′) *there is a decreasing sequence converging to any given point x of* $\mathfrak{R}$.

For Σ^1 being a covering there is a $U^{1\alpha_1}\supset x$. Owing to (a) then and to the regularity of $\mathfrak{R}$, there is a $U^{n_2\alpha_2}$ such that $x\subset U^{n_2\alpha_2}\Subset U^{1\alpha_1}$, then a $U^{n_3\alpha_3}$ such that $x\subset U^{n_3\alpha_3}\Subset U^{n_2\alpha_2}$, etc. Therefore $\{U^{n_i\alpha_i}\}$ is a decreasing sequence all of whose sets $\supset x$, hence by (d) it converges to x.

Conversely suppose that (a′), (b), (c), (d) hold for Δ. Whatever the open set $V\supset x$, there is a $U^{n_i\alpha_i}\supset x$ and $\subset V$, hence V is a sum of U's and (a) holds.

Another noteworthy property is that both (a) and (d) are consequences of

(e) *the mesh of* $\Sigma^i\to 0$ with $1/i$.

For let it be ε_i and let $\Pi U^{n_i\alpha_i} \supset x+y$, $x \neq y$. For a certain i we will have diam $U^{n_i\alpha_i} < d(x, y)$, and hence $U^{n_i\alpha_i} \not\supset x+y$, i. e. the U's cannot converge to more than one point. As regards (a), let V be open and $\supset x$. Since $\mathfrak{R}-V$ is closed, $d(x, \mathfrak{R}-V) = \varepsilon > 0$. For some i mesh $\Sigma^i < \varepsilon$. Since Σ^i is a covering there will be an $U^{i\alpha} \supset x$ and hence $\subset V$ since diam $U^{i\alpha} < \varepsilon$. Therefore V is a sum of sets $U^{i\alpha}$, which shows that (a) follows likewise from (e).

6. **Existence theorem.** *Every separable metric space possesses a normal development.* Before we take up the formal proof we will make an observation as regards Lemma I. Suppose that we have a first finite aggregate of non-intersecting pairs of closed sets (A^i, B^i), $i = 1, 2, \cdots, p_1$, then increase it by another (A^{p_1+i}, B^{p_1+i}), $i = 1, 2, \cdots, p_2$, etc. We can construct the characteristic functions $g_i(x)$, for the first p_1 sets, then the new functions $g_{p_1+i}(x)$ arising for the second set, etc. Consider now the set R represented by (2) as before and let R^i denote the locus obtained when in (2) all functions $g(x)$ beyond g_{p_i} are replaced by zero. It has been shown in the course of the proof of Lemma I that the R's are all homeomorphs of $\mathfrak{R}$, and R^i is the R of the lemma corresponding to the first p_i pairs. Now let $d_i(x, x')$ denote the distance-function for $\mathfrak{R}$ corresponding to the metric of R^i: distance equal to that of the images of the points x, x' on R^i. By writing down the expressions of the distances, we verify immediately that $d_{i+1}(x, x') \geqq d_i(x, x')$. Hence *for fixed* x, x', $d_i(x, x')$ *is a non-decreasing function of* i. This is the property which we need in the immediate sequel.

7. To prove our theorem we must construct the successive Σ's of Δ. We take for Σ^1 a wholly arbitrary f. c. o. s. and choose a metric, i. e. a distance-function $d_1(x, y)$, such that Lemma II becomes applicable to Σ^1, the η of the lemma being denoted by η_1. We then choose for Σ^2 a f. c. o. s. whose mesh $\varepsilon_2 < \eta_1$, so that we will have $\Sigma^2 \in \Sigma^1$. Σ^2 may be determined thus: $d_1(x, y)$ is the actual distance-function of a certain topological image R of $\mathfrak{R}$ on $\mathfrak{H}$. Since $\mathfrak{H}$ is compact it possesses an ε_2-f. c. o. s., and the sets of the latter intersect R in sets which are open, of diameter $< \varepsilon_2$, and hence their images on $\mathfrak{R}$ are the sets of a suitable Σ^2.

We now impose a second modification upon our metric, that is choose a new distance function $d_2(x, y)$, as outlined in No. 6, such that: (a) Lemma II becomes applicable to Σ^2, thus giving rise to an η_2 for it; (b) whenever $U^{2\alpha} \in U^{1\beta}$ and hence $\bar{U}^{2\alpha} \cdot A^{1\beta} = 0$, $A^{i\beta} = \mathfrak{R} - U^{i\beta}$, we have also $d_2(\bar{U}^{2\alpha}, A^{1\beta}) > 0$. The distances mentioned in (b) being in finite number they will have a lower bound $\xi_2 > 0$. The passage from d_1 to d_2 is by augmenting the functions g_i after the manner of No. 6, so that we will

have $d_2(x, y) \geqq d_1(x, y)$ whatever $x, y \subset \mathfrak{R}$. We now choose for Σ^3 any f. c. o. s. whose mesh $\varepsilon_3 < \xi_2$ or η_2. Suppose that we have thus reached Σ^i. We choose $d_i(x, y)$ as outlined in No. 6, such that: (a) Lemma II holds for Σ^i with a corresponding η_i; (b) whenever $U^{i\alpha} \Subset U^{j\beta}$, $j < i$, and hence $\bar{U}^{i\alpha} \cdot A^{j\beta} = 0$, also $d_i(\bar{U}^{i\alpha}, A^{j\beta}) > 0$. There will be again a positive lower bound ξ_i for the latter distances, and we choose Σ^{i+1} of mesh $\varepsilon_{i+1} < \xi_i$ or η_i. We do this for every i with the added restriction that $\varepsilon_i \to 0$ monotonely with increasing i. Since $\varepsilon_{i+1} < \eta_i$ we have $\Sigma^{i+1} \Subset \Sigma^i$ for every i. From $\varepsilon_{i+k} \leqq \varepsilon_{i+1} < \xi_i$ follows that every $\bar{U}^{i+k,\beta}$ which meets $\bar{U}^{i\alpha}$ is on any $U^{j\gamma} \Supset U^{i\alpha}$, $j < i$, and this for every i. Therefore conditions (b), (c) for a normal development hold throughout $\Delta = \{\Sigma^i\}$.

The choice of our metric is such that $d_{i+1}(x, y) \geqq d_i(x, y) \geqq d_1(x, y)$. Therefore the mesh of Σ^i, as measured by the fixed distance-function $d_1(x, y)$, $\to 0$ monotonely like ε_i. From this follows that conditions (a), (d) for a normal Δ hold also (No. 5). Therefore Δ is normal.

REMARK. Referring to the lemmas we find that when $\mathfrak{R}$ is compact there is no need to modify the metric. Therefore

THEOREM. *When $\mathfrak{R}$ is compact any development Δ consisting of a sequence of f. c. o. s. is normal provided that mesh Σ^i tends rapidly enough to zero with $1/i$.*

8. **Criterion for compactness.** *$\mathfrak{R}$ is compact when and only when every decreasing sequence $\{U^{n_i\alpha_i}\}$ converges to a point.*

When $\mathfrak{R}$ is compact $\Pi U^{n_i\alpha_i} = \Pi \bar{U}^{n_i\alpha_i} \neq 0$, and hence it consists of a point.

Conversely suppose that the intersection is always a point. Let A be an infinite subset of $\mathfrak{R}$, and let K^1 be the aggregate of sets $U^{1\alpha}$ each of which carries an infinite number of points of A. Since A is infinite, $K^1 \neq 0$. In the class K^1 there is a subclass K^2 consisting of sets $U^{1\alpha}$ such that $U^{1\alpha} \Supset U^{2\beta}$, where $U^{2\beta}$ carries an infinite number of points of A. Similarly K^2 has a subclass K^3 consisting of all sets $U^{1\alpha} \Supset U^{2\beta} \Supset U^{3\beta}$, where $U^{3\beta}$ carries an infinite number of points of A, etc. Whatever i, there is a $U^{i\alpha_i}$ with an infinite number of points of A. By condition (b) there is a $U^{i-1,\alpha_{i-1}} \Supset U^{i\alpha_i}$ and thus a sequence $U^{1\alpha_1} \Supset U^{2\alpha_2} \cdots \Supset U^{i\alpha_i}$. Therefore $K^i \neq 0$, and hence k_i, the number of sets $U^{1\alpha}$ in K^i, > 0. Also $K^i \subset K^{i-1}$. Hence $k_1 \geqq k_2 \geqq \cdots > 0$. Therefore the sequence k_1, $k_2, \cdots$ has a least positive integer say k_p. The sets of the class K^p, will be sets $U^{1\alpha}$, such that there corresponds to each a decreasing infinite sequence $\{U^{i\alpha_i}\}$ on each of whose sets there is an infinite number of points of A. We have by assumption $\Pi U^{i\alpha_i} = x$, a point, and owing to condition (c), every neighborhood of x contains a set $U^{i\alpha_i}$, hence an infinite

number of points of A. The point x is then a condensation-point of the arbitrary infinite subset A of $\mathfrak{R}$, and $\mathfrak{R}$ is therefore compact.

9. With any normal development Δ there is associated a mapping of $\mathfrak{R}$ on a certain compact space which is basic in the next section. Consider the following sets of non-intersecting pairs of closed sets: $(\bar{U}^{i\alpha}\,\bar{U}^{j\beta})$; $(\bar{U}^{i\alpha_0}\ldots\ldots\bar{U}^{i\alpha_p}, \bar{U}^{i\beta})$; $(\bar{U}^{i\alpha}, A^{j\beta})$, where $A^{j\beta} = \mathfrak{R} - U^{j\beta}$. They include the pairs of Urysohn's mapping described in No. 2. If $f_i(x)$, $i = 1, 2, \cdots + \infty$, are characteristic functions corresponding to all these pairs, (1) will again represent a topological image R of $\mathfrak{R}$ on $\mathfrak{H}$. If $U^{i\alpha}$, $A^{j\beta}$ continue to designate the images of the sets thus called previously, we will have now that whenever two closed sets in a pair such as above do not meet their distance on R is >0.

10. Let R^* designate the closure of R on $\mathfrak{H}$. Since the latter is compact, R^* is self-compact. We have thus mapped $\mathfrak{R}$ topologically on a set R^* which is compact as a space. This is the mapping that we have been seeking.

Now the closure $A^{*i\alpha}$ of $A^{i\alpha}$ on $\mathfrak{H}$ is closed also relatively to R^*. Hence $U^{*i\alpha} = R^* - A^{*i\alpha}$ is relatively open. The points of R on $U^{*i\alpha}$ are all its points at a distance >0 from $A^{*i\alpha}$, hence also from $A^{i\alpha}$, and their sum is $U^{i\alpha}$, so that $U^{i\alpha} \subset U^{*i\alpha}$. Moreover if x is a condensation-point of R on $U^{*i\alpha}$, the points of R very near to x are at a distance >0 from $A^{i\alpha}$, and hence again $\subset U^{i\alpha}$, so that $U^{*i\alpha} = U^{i\alpha} +$ condensation-points of $U^{i\alpha}$ on $\mathfrak{H}$. *Therefore the closures of $U^{i\alpha}$ and $U^{*i\alpha}$ relatively to $\mathfrak{H}$ coincide.*

Let the closure bars refer respectively to closures on R or R^* as regards their subsets. Since R^* is closed the closures relatively to it and to $\mathfrak{H}$ are the same. It follows that $\bar{U}^{*i\alpha}$ is the closure of $U^{i\alpha}$ on $\mathfrak{H}$.

From the closure relations between the U's and U^*'s, we deduce:

(a) $\bar{U}^{*i\alpha} \cdot \bar{U}^{*j\beta} \neq 0$,
(b) $U^{*i\alpha} \in U^{*j\beta}$,
(c) $\bar{U}^{*i\alpha_0}\ldots\ldots\bar{U}^{*i\alpha_p} \neq 0$,

} are equivalent to the same non-starred relations;

(d) mesh $\Sigma^{*i} \to 0$ with $1/i$, $\Sigma^{*i} = \{U^{*i\alpha}\}$.

The last property is evident. From our basic closure relations between the U's follows that

$$d(\bar{U}^{i\alpha}, \bar{U}^{j\beta}) = d(\bar{U}^{*i\alpha}, \bar{U}^{*j\beta}), \tag{3}$$

$$d(\bar{U}^{i\alpha}, A^{j\beta}) = d(\bar{U}^{*i\alpha}, A^{*j\beta}). \tag{4}$$

Now to impose the existence of one or the other intersection in (a) is equivalent to demanding that the distances (3) vanish, which proves (a). Similarly (b) is a consequence of the fact that to impose one of the inclusions in it is equivalent to demanding that the distances in (4) be

>0. As regards (c) since the closures as to R^* are also closures as to $\mathfrak{H}$, we have

$$\bar{U}^{i\alpha_0} \cdot \ldots \cdot \bar{U}^{i\alpha_p} = \bar{U}^{*i\alpha_0} \cdot \ldots \cdot \bar{U}^{*i\alpha_p} \cdot R.$$

Therefore when the sets $\bar{U}^{i\alpha}$ intersect so do the sets $\bar{U}^{*i\alpha}$. Conversely if the intersection of the non-starred sets $= 0$, there is a $q<p$ such that $\bar{U}^{i\alpha_0} \cdot \ldots \cdot \bar{U}^{i\alpha_q} \neq 0$, $\bar{U}^{i\alpha_0} \cdot \ldots \cdot \bar{U}^{i\alpha_{q+1}} = 0$. Therefore by our construction of $\mathfrak{R}$ (see No. 9)

$$d(\bar{U}^{i\alpha_0} \cdot \ldots \cdot \bar{U}^{i\alpha_q},\ \bar{U}^{i\alpha_{q+1}}) > 0,$$

and since the distances are the same as for the corresponding starred sets, necessarily

$$\bar{U}^{*i\alpha_0} \cdot \ldots \cdot \bar{U}^{*i\alpha_{q+1}} = 0 \cdots = \bar{U}^{*i\alpha_0} \cdot \ldots \cdot \bar{U}^{*i\alpha_p}.$$

From (a), (b), (d) we conclude that conditions (a′), (b), (c), (d) of No. 5 for a normal development hold as regards $\Delta^* = \{\Sigma^{*i}\}$.

The set R is clearly dense on R^*. Conversely if R^* is self-compact, with Δ^* as a normal development, and R is a subset of R^* which is dense on the latter, the sets $U^{i\alpha} = U^{*i\alpha} \cdot R$ define on R a normal development Δ related to Δ^* in the same manner as above. We have therefore proved:

THEOREM. *To every normal development Δ of $\mathfrak{R}$ there corresponds a dense topological mapping of $\mathfrak{R}$ on a compact metric space $\mathfrak{R}^*$, on which there exists a normal development Δ^* such that the i-th coverings of each have the same intersection properties (expressed by condition* (c)*). Conversely given Δ^* on $\mathfrak{R}^*$, there is defined on any dense subset R of $\mathfrak{R}^*$ a normal development Δ related to Δ^* in the preceding manner.*

§ 3. THE ORDER OF A SPACE.

11. The *order* of an aggregate of sets $\sigma = \{A^\alpha\}$, is the largest integer $\varrho(\sigma)$ such that there are $1+\varrho(\sigma)$ sets A with a non-vacuous intersection.

Let Σ be a fixed f. c. o. s., σ a variable one $\subset \Sigma$. The number $\varrho(\sigma)$ being a positive integer has a least value $\omega(\Sigma)$ which is actually reached for some σ. As Σ itself varies $\omega(\Sigma)$ has a highest value finite or infinite, ω, called the *order* of $\mathfrak{R}$. When ω is finite it is reached for some Σ. In any case ω is a topological invariant of $\mathfrak{R}$.

Instead of a variable f. c. o. s. σ, we might have coverings by other types of sets. Any topological class $\mathfrak{K}$ of such sets gives rise to a topological invariant $\omega(\mathfrak{R}, \mathfrak{K})$ of $\mathfrak{R}$ analogous to ω. The following two classes are noteworthy:

(a) $\mathfrak{K}^1$, the class of all closures of open sets.

(b) $\mathfrak{K}^2$, the class of all closed sets.

The second class is the one selected by all previous writers. Let ω_1, ω_2 be the ω's corresponding to the two new classes. I say that $\omega_1 = \omega_2 = \omega$.

12. Since $\mathfrak{R}^1 \subset \mathfrak{R}^2$ it follows immediately from the definition of the ω's that $\omega_2 \leqq \omega_1$. Take now two f. c. o. s., σ, Σ, with $\sigma \subset \Sigma$. As recalled in the proof of Lemma II, according to Menger there can be found, for every α, a $V^\alpha \in U^\alpha$ such that $\sigma' = \{V^\alpha\}$ is likewise a f. c. o. s. Let $\overline{\sigma}' = \{\overline{V}^\alpha\}$. We manifestly have $\varrho(\overline{\sigma}') \leqq \varrho(\sigma)$, hence the lower bound of $\varrho(\overline{\sigma}') \leqq \omega(\Sigma)$. As the sets of $\overline{\sigma}'$ belong to $\mathfrak{R}^1$, passing to the upper bounds of the two orders we find $\omega_1 \leqq \omega$.

13. We will now show that $\omega_2 \geqq \omega$. Let $\sigma = \{A^\alpha\}$ be a finite covering by closed sets $\subset \Sigma = \{U^\beta\}$. By Lemma II the metric of $\mathfrak{R}$ can be so chosen that the distances $d(A^\alpha, \mathfrak{R} - U^\beta)$ corresponding to all pairs A^α, U^β such that $A^\alpha \subset U^\beta$, have a lower bound $\xi_1 > 0$. By Lemma III it can be so chosen that there exists a $\xi_2 > 0$ such that, if $A^{\alpha_0} \ldots . A^{\alpha_p} \neq 0$, at least one distance $d(x, A^{\alpha_i}) > 0$ whatever the point x of $\mathfrak{R}$. The choice of the two metrics involves a mapping of $\mathfrak{R}$ on $\mathfrak{H}$ by a system (2) with suitable functions $g_i(x)$. The two sets of functions are finite. If we unite them into a single set and take the representation (2) corresponding to that set, we find that both properties above mentioned are fulfilled simultaneously by the metric associated with the new representation, and we have as above a ξ_1 and a ξ_2; let ξ be the least of the two.

Let now V^α be the set of all points of $\mathfrak{R}$ whose distance from $A^\alpha < \xi$. It is an open set, and if $A^\alpha \subset U^\beta$ we have $V^\alpha \in U^\beta$. Hence

$$\sigma \subset \sigma' = \{\overline{V^\alpha}\} \in \{U^\beta\}$$

and σ' is a f. c. o. s. Moreover $V^{\alpha_0} \ldots . V^{\alpha_p} \neq 0$ when and only when $A^{\alpha_0} \ldots . A^{\alpha_p} \neq 0$. That the second implies the first is clear. Conversely let the first hold, and the second not. Then a point of $V^{\alpha_0} \ldots . V^{\alpha_p}$ is nearer than $\xi \leqq \xi_2$ to the sets A^{α_i}, which contradicts the basic property of ξ_2. Thus when one of the two intersections exists so does the others. It follows that

$$\varrho(\sigma) = \varrho(\sigma').$$

From this we conclude, as in No. 12, that

$$\omega(\Sigma) \leqq \varrho(\sigma),$$

hence passing to the maxima, $\omega_2 \geqq \omega$. Since $\omega_2 \leqq \omega_1 \leqq \omega$, these three numbers are equal.

14. **Order of a normal development.** If $\Delta = \{\Sigma^i\}$ little can be said in general about $\varrho(\Sigma^i)$. Suppose however that Δ has the following property: for every i,

$$\varrho(\Sigma^{i+1}) = \varrho(\bar{\Sigma}^{i+1}) = \omega(\Sigma^i),$$

its least possible value consistent with $\Sigma^{i+1} \in \Sigma^i$. Owing to this inclusion $\varrho(\Sigma^{i+1}) \geqq \varrho(\Sigma^i)$, and hence $\varrho(\Sigma^i)$ is a non-decreasing function. Its limit, Ω, finite or infinite, is called the order of Δ, and a development of the type considered is said to be *minimal*. Since

$$\varrho(\Sigma^{i+1}) = \omega(\Sigma^i),$$

necessarily $\omega \geqq \Omega$.

THEOREM. *A space* $\mathfrak{R}$ *of finite order* ω *possesses minimal normal developments whose order is also* ω.

We obtain Δ such as required by a suitable modification of the construction in § 2. With the same notations as loc. cit., we take, as we may by No. 13, Σ^1 such that

$$\varrho(\Sigma^1) = \varrho(\bar{\Sigma}^1) = \omega,$$

its maximum value. Suppose that we have obtained $\Sigma^1, \cdots, \Sigma^{i-1}$, such that

$$\varrho(\Sigma^{i-1}) = \varrho(\bar{\Sigma}^{i-1}) = \omega.$$

We now take an ε_i-f. c. o. s., σ^i, hence $\in \Sigma^{i-1}$, and choose $\Sigma^i \subset \sigma^i$, such that

$$\varrho(\Sigma^i) = \varrho(\bar{\Sigma}^i) = \omega(\sigma^i).$$

We have then

$$\omega(\sigma^i) \geqq \omega(\Sigma^{i-1}) = \omega, \quad \text{and} \quad \omega(\sigma^i) \leqq \omega.$$

Hence

$$\varrho(\Sigma^i) = \omega(\sigma^i) = \omega.$$

For the usual reason likewise

$$\omega(\Sigma^i) \geqq \omega(\sigma^i) = \omega,$$

and also $\leqq \omega$, hence $\omega(\Sigma^i) = \omega$. By this construction then

$$\varrho(\Sigma^i) = \varrho(\bar{\Sigma}^i) = \omega(\Sigma^i) = \omega$$

for every $i > 1$, and hence Δ is normal and minimal of order ω.

15. **Order of a compact space.** When $\mathfrak{R}$ is compact an order may be defined in somewhat different manner. Take all ε-f. c. o. s., and let $r(\varepsilon)$ be their lowest order. Since an ε'-covering is also an ε-covering when $\varepsilon' < \varepsilon$, we have then $r(\varepsilon') \geqq r(\varepsilon)$. Hence $r(\varepsilon)$ is a non-decreasing function of ε and it has a maximum ω_0, finite or infinite, called the *metric* order. I say again that $\omega_0 = \omega$, *or the metric and topological orders are equal.*

Let first ω_0 be finite and construct Δ as follows: Σ^1 is of mesh ε_1 such that already $r(\varepsilon_1) = \omega_0$; Σ^i is taken of mesh $\varepsilon_i < \varepsilon_{i-1}$ or η_{i-1}, the η of Lemma II corresponding to Σ^{i-1}; we impose $\varrho(\Sigma^i) = \omega_0$, its minimum value since $\varepsilon_i < \varepsilon_1$, and also $\varepsilon_i \to 0$ with $1/i$. By Lemma II the sets of Σ^i can be slightly shrunk without its ceasing to be a covering, say Σ'^i. Again

$$\varrho(\Sigma'^i) = \omega_0,$$

and moreover

$$\varrho(\Sigma'^i) \leqq \varrho(\bar{\Sigma}'^i) \leqq \varrho(\Sigma^i) = \omega_0,$$

hence

$$\varrho(\bar{\Sigma}'^i) = \omega_0.$$

Writing now Σ^i for Σ'^i, we see that the development Δ will be minimal of order ω_0. Now when ω is finite it is precisely such a Δ that we have constructed in No. 15, therefore when ω, ω_0 are finite $\omega_0 = \omega$. When $\omega = \infty$ we have, whatever the integer n, a f. c. o. s., Σ whose $\omega(\Sigma) > n$. If η is the number of Lemma II for Σ, any ε-f. c. o. s. $\sigma \subset \Sigma$, whatever $\varepsilon < \eta$, hence its order

$$\varrho(\sigma) \geqq \omega(\Sigma) > n,$$

and therefore $r(\varepsilon) > n$ so that limit $r(\varepsilon) = \infty = \omega_0 = \omega$.

If in place of open sets we choose ε-coverings by closed sets we obtain a new metric order ω_0'. This is the order as defined by Urysohn. It is not difficult to show that it is again equal to ω. For let $\sigma = \{A^\alpha\}$ be a finite ε-covering by closed sets. By Lemma III we can shlightly dilate the A's without changing $\varrho(\sigma)$.—A^α is thus replaced by a certain $\bar{V}^\alpha$ and $\sigma' = \{V^\alpha\}$ is an ε'-f. c. o. s., where $\varepsilon' \to 0$ with ε, whose order is also $\varrho(\sigma)$. Hence $r(\varepsilon') \leqq \varrho(\sigma)$ and therefore $\omega_0 \leqq \omega_0'$.—On the other hand if we have an ε-f. c. o. s. σ, Lemma II authorizes us to contract its sets into a certain covering $\sigma' = \{V^\alpha\} \Subset \sigma$, and $\varrho(\bar{\sigma}') \leqq \varrho(\sigma)$. Since $\bar{\sigma}'$ is a special finite ε-covering by closed sets, we find this time $\omega_0 \geqq \omega_0'$, hence $\omega_0' = \omega_0 = \omega$.

16. **The Lebesgue order theorem.** *The order of* $\mathfrak{R}$ *is equal to its dimension.* This fundamental theorem was first stated for Euclidean spaces by Lebesgue, then proved for them by Brouwer, later also by Lebesgue. The complete theorem was proved for compact spaces by Urysohn,[5] while for separable spaces Menger[6] showed that $\omega \leqq n = \dim \mathfrak{R}$. By a judicious use of normal developments Urysohn's proof for the compact case can be made to yield the proof of the theorem for separable spaces.

We have to show that when $\mathfrak{R}$ is separable $\omega \leqq n$, and at the same time $\geqq n$. We indicate briefly the mild departures from Urysohn's proof.

[5] Fundamenta Matematicae, vol. 8 (1925), p. 292.

[6] *Dimensionstheorie*, Ch. V.

Let Σ be a given f. c. o. s. To prove $\omega \leqq n$ we have to find a f. c. o. s. $\sigma \subset \Sigma$ whose sets have boundaries that are at most $(n-1)$-dimensional. According to No. 7 we can find a f. c. o. s. $\sigma' = \{V^\beta\} \in \Sigma = \{U^\alpha\}$. Given V^β there exists a $U^\alpha \supset V^\beta$. Since the two closed sets $\bar{V}^\beta$, $\mathfrak{R} - U^\alpha$ do not meet there exists an open set W^β such that $V^\beta \in W^\beta \in U^\alpha$, $\dim F(W^\beta) < n$.[7] Therefore $\sigma = \{W^\beta\}$ is a f. c. o. s. behaving as required. From this point on the proof continues substantially as with Urysohn.

We now pass to $n \leqq \omega$. By Urysohn's method and a simple induction we can establish this result: if Δ is a normal development having the property that $\varrho(\Sigma^i) = \varrho(\bar{\Sigma}^i) = k$, a fixed integer, then $n \leqq k$. Since there are minimal developments of order ω, it follows that for them $k = \omega$, hence $n \leqq \omega$. This is then in outline the proof of the order theorem as it may now be derived by the aid of normal developments.

17. **The Imbedding theorem for a separable metric space:** $\mathfrak{R}$ *can be mapped topologically on a compact metric space of the same dimension.* We have seen in no. 10 that with any normal developmnt Δ of $\mathfrak{R}$ there is associated a topological mapping of $\mathfrak{R}$ on a compact metric space $\mathfrak{R}^*$ having the following property: In the notations loc. cit. there is a normal development Δ^* of $\mathfrak{R}^*$ such that the orders of Σ^i and Σ^{*i} are the same (no. 10 property c).

Suppose in particular that Δ is minimal of order $\omega = n$. Then

$$\varrho(\Sigma^i) = \varrho(\Sigma^{*i}) = n.$$

Since mesh $\Sigma^{*i} \to 0$ (No. 10 d), order $\mathfrak{R}^* = \dim \mathfrak{R}^* \leqq n$. But as $\mathfrak{R}^*$ contains an n-dimensional set, $\dim \mathfrak{R}^* \geqq n$, hence $\dim \mathfrak{R}^* = n$, which proves our theorem.

18. REMARK. Combined proofs of the order and imbedding theorems can be obtained along the following lines: (a) $\omega \leqq n$ is to be established as by Menger; (b) $\omega = n$ is to be proved as by Urysohn for the compact case only; (c) it is then proved as above that $\mathfrak{R}$ can be mapped topologically on a compact metric space $\mathfrak{R}^*$ whose order $\leqq \omega$. By (b) we then have $\dim \mathfrak{R}^* \leqq \omega$, and since $n \leqq \dim \mathfrak{R}^*$, this yields $n \leqq \omega$, and hence $n = \omega$, which is the Lebesgue order theorem. We then have $\dim \mathfrak{R}^* \leqq n$ and hence $= n$, which together with (c) yields the imbedding theorem.

19. COROLLARY. *The various immersion theorems obtained in our paper already quoted (footnote 2) are valid for any separable metric space. In particular the two universal n-spaces there considered are also universal n-spaces for separable metric n-spaces.*

[7] Menger, loc. cit., p. 115.

20. **Menger's generalization of the Lebesgue order theorem.** Let Σ be a f. c. o. s., $\sigma = \{A^\alpha\}$ a finite covering by closed sets $\subset \Sigma$. Instead of the order $\varrho(\sigma)$ we may consider with Menger the maximum dimension $\varrho_k(\sigma)$ of any intersection $A^{\alpha_0} \cdot \cdots \cdot A^{\alpha_k}$. By the same considerations as before we obtain an invariant d_k analogous to the order, and we then have, with $\dim \mathfrak{R} = n$:

THEOREM (Menger). $d_k = n-k$.

For $k = n+1$ this yields $d_{n+1} = -1$, which is merely the expression of the Lebesgue order theorem. As in the proof of the latter we have to show that $d_k \leqq n-k$ and also $\geqq n-k$. The first part was proved by Menger, the second also by him for a compact space, (loc. cit. p. 156 and following). Thus all that is now lacking is the inequality $d_k \geqq n-k$ for any separable metric space.

21. By repetition of the argument used in proving the theorem in No. 14 we show that there exists a development Δ such that $\varrho_k(\bar{\Sigma}^i) = d_k$ for every i. To complete the proof there remains to be shown as in the similar case in No. 16, that if there exists a Δ such that $\varrho_k(\bar{\Sigma}^i) \leqq \delta$ for every i, then $n \leqq \delta+k$; since δ can take the value d_k, the required inequality will follow. We have already disposed of $\delta = 0$ (No. 16), hence we can use induction on δ.

Suppose then that $\varrho_k(\bar{\Sigma}^i) \leqq \delta$ for every k. Now all the sets $\bar{U}^{i\alpha_0} \cdot \cdots \cdot \bar{U}^{i\alpha_k} \neq 0$ constitute an enumerable aggregate of closed sets. Let us designate them, ranged in some order by $\{G^\beta\}$, $\beta = 1, 2, \cdots, +\infty$. Since $\dim G^\beta \leqq \delta$, by a well known theorem[8] $\dim \Sigma G^\beta \leqq \delta$. It follows that any point x of $\mathfrak{R}$ possesses a neighborhood V whose boundary intersects the sum, and hence every G, in a set whose dimension $\leqq \delta-1$[9]. But the intersections $U^{i\alpha} \cdot F(V)$ determine on $F(V)$ a normal development such that the corresponding G's are the sets $G^\beta \cdot F(V)$, and these sets are $\delta-1$ dimensional at most. Under the hypothesis of the induction then $\dim F(V) \leqq \delta-1+k$, and hence $\dim \mathfrak{R} = n \leqq \delta+k$. This completes the proof of Menger's theorem.

[8] Hurewicz, Math. Ann., vol. 96 (1927), p. 760. See also Menger, loc. cit. p. 96.

[9] Menger, loc. cit. p. 82.

PRINCETON, N. J.

(ADDED IN THE PROOF.) Professor Čech recently called our attention to a paper by Hurewicz, Monatshefte für Math. und Physik, 37 (1930), pp. 199–208, which had escaped us, and where he proves this imbedding theorem by a procedure based like ours on a modification of metric. While we have many points of contact with Hurewicz we diverge in scope, and more particularly in the explicit consideration of normal developments and in the treatment of the order.

A CLASSIFICATION OF THE HOMOLOGY GROUPS OF COMPACT SPACES.[1]

By J. W. Alexander and L. W. Cohen.

1. The homology groups of a compact space have been defined by Vietoris.[2] We give below a classification of these groups by means of invariants. The problem will be treated from the point of view of abstract group theory. In the applications to analysis situs the group G of § 2 will be the group of all i-cycles of a compact space, and the closed subgroup H of § 8 will be the group of all i-cycles that are homologous to zero.

A. The infinite commutative group G.

2. Let

$$x_1, x_2, x_3, \cdots$$

be an infinite sequence of marks and let

$$p = \sum_s a_s x_s$$

be any infinite linear form in these marks, with integer coefficients. Then, to every form p there corresponds a *negative,*

$$-p = \sum_s -a_s x_s,$$

and to every pair of forms p and p' there corresponds a *sum,*

$$p+p' = \sum_s (a_s + a_s') x_s.$$

The set of all forms (p) may, therefore, be regarded as a group G with respect to the operation of addition.

If the first $n-1$ coefficients of a form p are zero, but not the n-th, we shall call the reciprocal of the number n the *norm* $\omega(p)$ of p,

$$\omega(p) = \frac{1}{n}.$$

---

[1] Received March 7, 1931. Part of a paper read before the Amer. Math. Soc., April 3, 1931, under the title "Homology Groups Associated with Infinite Complexes".

[2] L. Vietoris. "Über den höheren Zusammenhang kompakter Räume ...", Math. Ann. 97 (1926), pp. 454–472.

If all the coefficients of p vanish we shall set the norm of p equal to zero,

$$\omega(0) = 0.$$

Clearly, the norm of any multiple of p is equal to the norm of p itself,

$$\omega(\lambda p) = \omega(p), \qquad \lambda \neq 0. \tag{1}$$

Moreover, the relation

$$\omega(p) > \omega(q)$$

implies the relation

$$\omega(p+\lambda q) = \omega(p). \tag{2}$$

3. We shall think of the forms p as the *points* of a *group space* Σ_G and shall define the distance $\delta(p, q)$ between two points p and q of Σ_G as the norm of the difference $(p-q)$,

$$\delta(p, q) = \omega(p-q).$$

The relation of symmetry,

$$\delta(p, q) = \delta(q, p), \tag{3}$$

follows at once from (1) by setting $\lambda = -1$.

The distance function δ has the following property:

Let p, q, and r be any three points of the group space Σ_G. Then either all three of the distances $\delta(p, q)$, $\delta(p, r)$, and $\delta(q, r)$ are equal or else two of them are equal and the third one is smaller than the other two.

For if the three distances are not all equal, we may choose our notations so as to have

$$\delta(p, q) > \delta(q, r),$$

or, in other words,

$$\omega(p-q) > \omega(q-r).$$

We shall then have

$$\begin{aligned} \delta(p, r) = \omega(p-r) &= \omega[(p-q)+(q-r)] \\ &= \omega(p-q), \qquad \text{by (2)}, \\ &= \delta(p, q), \end{aligned} \tag{4}$$

which is what we set out to prove.

It follows, at once, that the distance function satisfies the triangle inequality,

$$\delta(p, q) \geqq \delta(p, r)+\delta(q, r), \tag{5}$$

for arbitrary choices of the points p, q, and r. *The space Σ_G is, therefore, a metric space.*

4. Let us next inquire under what conditions an infinite sum of linear forms,

$$p_1 + p_2 + p_3 + \cdots,$$

converges to a limit. Suppose we have

$$p_i = \sum_s a_{is} x_s, \qquad (i = 1, 2, 3, \cdots). \tag{6}$$

Then, *a necessary and sufficient condition that the infinite sum* $\sum p_i$ *converge to a limit is that the norms of the forms* p_i *converge to zero,*

$$\lim_{i \to \infty} \omega(p_i) = 0. \tag{7}$$

Condition (7) is obviously necessary. It is also sufficient. For if it is fulfilled there can be at most a finite number of non vanishing elements a_{ij} in each column of the matrix $\| a_{ij} \|$. Let a_s be the sum of all the elements in the s'th column of $\| a_{ij} \|$ and let p be the form

$$p = \sum_s a_s x_s.$$

Then, clearly, the norm of the difference

$$p - (p_1 + p_2 + \cdots + p_n)$$

approaches zero as n increases indefinitely. This establishes the sufficiency of the condition.

Condition (7) also implies

$$\lim_{i \to \infty} \omega(b_i p_i) = 0,$$

where the b_i's are arbitrary integers. Therefore, *if condition* (7) *is fulfilled, every linear combination of the forms* p_i *converges to a limit,*

$$\sum_i b_i p_i = \sum_{i,s} b_i a_{is} x_s. \tag{8}$$

The coefficients $\sum_i b_i a_{is}$ in the right hand member of (8) are, of course, finite sums because there are at most a finite number of non vanishing elements a_{is} for each value of s.

A matrix $\| a_{ij} \|$ will be said to be *regular* if there are at most a finite number of non vanishing elements in each of its columns. Condition (7) is clearly equivalent to the condition that the matrix $\| a_{is} \|$ in (6) be regular.

Let us note for future reference the following simple theorem:

The product of two regular matrices $\|a_{is}\|$ *and* $\|b_{sj}\|$ *is a well defined regular matrix.*

Proof. Since the matrix $\|b_{sj}\|$ is regular each term of the product matrix is a *finite* sum of the type

$$c_{ij} = \sum_s a_{is} b_{sj}. \tag{9}$$

The product matrix $\|c_{ij}\|$ is, therefore, well defined. To prove that $\|c_{ij}\|$ is regular let us assign an arbitrary fixed value to the index j. Then, because of the regularity of the matrix $\|b_{sj}\|$, there exists a number α such that in the jth column of $\|b_{sj}\|$ all elements after the one in the αth row are equal to zero. Moreover, because of the regularity of the matrix $\|a_{is}\|$ there exists a number β such that in the first α columns of $\|a_{is}\|$ all elements after the ones in the βth row are equal to zero. Therefore, as soon as i is greater than β the elements c_{ij} in (9) must all vanish. In other words there can be at most a finite number of non-vanishing elements in each column of the product matrix $\|c_{ij}\|$.

B. Change of basis.

5. Let us now consider a new infinite sequence of marks,

$$y_1, y_2, y_3, \cdots,$$

and construct the set of all linear combinations of these marks with integer coefficients. Let us also set up an arbitrary one-one correspondence between the elements p of the group G and the various linear combinations of the marks y_i such that:

(i) If

$$p = \sum_s a_s x_s$$

is any element of the group G and

$$q = \sum_s b_s x_s$$

the corresponding linear combination of the y_i's, then $-p$ always corresponds to $-q$.

(ii) If p and p' are any two elements of G and q and q' the two corresponding linear combinations of the y_i's then $p+p'$ always corresponds to $q+q'$.

Then, if we think of the elements p as functions of the y_i's rather than of the x_i's and if we redefine norm and distance in terms of the y_i's we shall, in general, alter the metric of the group space Σ_G. *If the topological*

structure of the space Σ_G remains invariant—that is to say, if the new metric leads to the same definition of limit points as the old—we shall say that the elements

$$y_i = \sum_s a_{is} x_s$$

of G form a *basis* of the group G.

A necessary and sufficient condition that a set of equations

$$y_i = \sum_s a_{is} x_s \tag{10}$$

determine a change of basis is that the matrix $\|a_{is}\|$ be regular and that it have a regular inverse $\|b_{is}\|$.

Proof. If the marks y_i are to be a basis of the group G every mark x_i must correspond to a linear combination of the marks y_i,

$$x_i = \sum_s b_{is} y_s. \tag{11}$$

Now, the matrix $\|b_{is}\|$ determined by expressions (11) must be regular. For the left-hand members of (11) converge to zero in terms of the metric of the x_i's; therefore, the right hand members must converge to zero in terms of the metric of the y_i's. The matrix $\|a_{is}\|$ in (10) must also be regular, for a similar reason. Thus, to prove the necessity of the condition we have only to show that the matrix $\|b_{is}\|$ is the inverse of the matrix $\|a_{is}\|$. To do this, let us form the following linear combination of equations (10),

$$\sum_s b_{is} y_s = \sum_{s,j} b_{is} a_{sj} x_j. \tag{12}$$

Then, by (11), the right hand member of (12) must, of necessity, converge to x_i, which means that we must have

$$\sum_s b_{is} a_{sj} = \delta_{ij}, \qquad \delta_{ij} = \begin{cases} 1 \text{ for } i = j \\ 0 \text{ for } i \neq j \end{cases}. \tag{13}$$

Moreover, by reversing the rôles of the two sets of equations (10) and (11) we may obtain the further relations

$$\sum_s a_{is} b_{sj} = \delta_{ij}. \tag{14}$$

Relations (13) and (14) signify that the matrix $\|b_{is}\|$ is the inverse of the matrix $\|a_{is}\|$.

To prove the sufficiency of the condition let us assume that the matrix $\|a_{is}\|$ is regular and that there exists a regular matrix $\|b_{is}\|$ satisfying condition (13) and (14). Then, with the aid of equations (10) we may

set each linear form in the y_i's into correspondence with a linear form in the x_i's according to the following rule:

$$\sum_i c_i\, y_i = \sum_{i,s} c_i\, a_{is}\, x_s,$$

while with the aid of equations (11) we may set each linear form in the x_i's into correspondence with a linear form in the y_i's:

$$\sum_i \alpha_i\, x_i = \sum_{i,s} \alpha_i\, b_{is}\, y_s.$$

These two correspondences must be identical. For the right hand member of the first correspondence is paired in the second correspondence with the form

$$\sum_{i,s,j} c_i\, a_{is}\, b_{sj}\, y_j = \sum_{i,j} c_i\, \delta_{ij}\, y_j = \sum_i c_i\, y_i, \quad \text{by (14)},$$

just as it is in the first. Similarly, the right hand member of the second correspondence is paired in the first correspondence with the form $\sum \alpha_i\, x_i$, (by (13)), just as it is in the second. It follows, at once that the correspondences are one-one, for the first relation gives the form in the x_i's corresponding to an arbitrary form in the y_i's, and the second relation gives the form in the y_i's corresponding to an arbitrary form in the x_i's.

To complete the proof of the sufficiency of the condition we must show that the metric of the y_i's leads to the same definition of convergence as the metric of the x_i's. Now, a sequence of elements p_i of the group G converges to a limit p if, and only if, the differences $p_i - p$ converge to zero. Let us write

$$p_i - p = \sum_s c_{is}\, x_s = \sum_{s,j} c_{is}\, b_{sj}\, y_j. \tag{15}$$

Then if the sequence $p_i - p$ converges to zero in the metric of the x_s's the matrix $\|c_{is}\|$ must be regular; therefore, the product matrix $\|\sum c_{is}\, b_{sj}\|$ must be regular, (§ 4); therefore the sequence $p_i - p$ must also converge to zero in the metric of the y_i's. A similar argument shows that if $p_i - p$ converges in the y_i's it also converges in the x_i's. This completes the proof.

The restriction that a transformation from one basis to another shall not alter the topological structure of the group space Σ_G seems to be required if we want the transformation to be completely determined by a set of relations of the form (10). Without such a restriction, relations (10) apparently tell us nothing about the form in the x_i's corresponding to an infinite linear combination of the y_i's, even when the matrix $\|a_{ij}\|$ is regular. It is barely possible, however, that every one-one correspondence between the set of all linear forms in the x_i's and the set of all linear forms in the y_i's such that the correspondence satisfies conditions (i)

and (ii) must automatically leave invariant the topological structure of the group space, and that the restrictions turn out to be redundant. At all events, we know of no example to the contrary.

6. The following two examples may, perhaps bring out more fully the significance of the theorem of § 5.

(i) If we write

$$\begin{aligned} y_1 &= x_1, \\ y_i &= -x_{i-1} + x_i, \qquad (i>1), \end{aligned} \tag{16}$$

we obtain, on solving for the x_i's, the inverse relations

$$x_i = y_1 + y_2 + \cdots + y_i. \tag{17}$$

Now, although the matrix of (16) is regular and has an inverse—namely the matrix of (17)—the inverse matrix is not regular; therefore, equations (16) do not determine a change of basis. In this particular example, the representations of the elements of G in terms of the y_i's is not unique. For instance, the sum of equations (16) converges to

$$y_1 + y_2 + y_3 + \cdots = 0.$$

(ii) The situation is somewhat different if we write

$$\begin{aligned} y_1 &= x_1, \\ y_i &= -2x_{i-1} + x_i, \qquad (i>1), \end{aligned} \tag{18}$$

in place of (16). The inverse of (18) is

$$x_i = 2^{i-1}y_1 + 2^{i-2}y_2 + \cdots + y_i, \tag{19}$$

therefore the inverse matrix is again irregular. In this example, no linear combinations of the y_i's can vanish identically. For a relation of the form

$$\sum_i c_i y_i = 0 \tag{20}$$

resulting from equation (18), can only hold if we have

$$\begin{aligned} c_{i+1} &= \frac{1}{2} c_i \\ &= \frac{1}{2^i} c_1. \end{aligned}$$

Moreover, the only integral solutions of these last expressions are $c_i = 0$, so that relation (20) must reduce to an identity. Thus, it is clear that no

two linear combinations of the y_i's can represent the same element of G. This time, however, not every element of G can be represented by a linear combination of the y_i's. Consider, for example, the linear combination

$$x_1 + x_3 + x_5 + \cdots. \tag{21}$$

If there were a linear combination of the y_i's reducing to (21) its coefficients would have to satisfy the relations

$$c_{i+1} = \frac{1}{2}(c_i - \varepsilon_i), \qquad \begin{cases} \varepsilon_i = 0, & i \text{ even}, \\ \varepsilon_i = 1, & i \text{ odd}. \end{cases}$$

Now, it is easy to verify that these relations admit no integral solutions. For if we assign any integral value to c_1 we find that the numerical values of successive coefficients must diminish steadily until one of them, say c_i, reaches one of the values: 0, 1, or -1, in which case either c_{i+1}, c_{i+2}, or c_{i+3} will have to be fractional.

We know of no example of a regular matrix $\|a_{ij}\|$ with an irregular inverse $\|b_{ij}\|$ such that the transformation

$$y_i = \sum_s a_{is} x_s$$

sets up a one-one correspondence between the set of all linear combinations of the y_i's and the set of all linear combinations of the x_i's. If such a transformation exists, the change from the metric of the x_i's to the metric of the y_i's will, of course, alter the topological structure of the group space Σ_G, in view of the theorem at the end of § 5.

Among the more obvious transformations which determine a change of basis are the following

(α) Any permutation of the marks x_i.

(β) Any transformation such that its matrix $\|a_{ij}\|$ has zeros below the main diagonal, 1's along the main diagonal, and arbitrary integer elements above. The inverse of $\|a_{ij}\|$ obviously exists and is of the same general type as $\|a_{ij}\|$ itself.

(γ) Any transformation determined by grouping the marks x_i into an infinite number of *finite* sets and performing on each finite set a transformation of determinant ± 1. Here again the inverse transformation will exist and will be obtainable by merely performing the inverse transformation on each finite set.

7. Consider an infinite sequence of transformations,

$$y_i(n) = \sum_s a_{is}(n)\, x_s \tag{22}$$

from a basis x_i to a variable basis $y_i(n)$, $(n = 1, 2, 3, \cdots)$. The sequence (22) will be said to *converge* to the limiting transformation

$$y_i = \sum_s a_{is} x_s \tag{23}$$

provided each of the coefficients $a_{is}(n)$ converges to the limit a_{is}; that is to say, provided each coefficient $a_{is}(n)$ differs from a_{is} for at most a finite number of values of n. If the transformations (22) converge to the limit (23) it does not necessarily follow that the limiting transformation determines a change of basis. Thus, for example, if the nth transformation (22) merely permutes the marks $x_1, x_2, \cdots, x_n$ cyclically the limiting transformation will be

$$y_i = x_{i+1},$$

and there will be no linear combination of the y_i's reducing to x_1. We, therefore, propose to find conditions under which a limiting transformation actually does determine a change of basis.

Let $\| a_{is}(n) \|$ be a variable regular matrix. Then the matrix $\| a_{is}(n) \|$ will be said to converge *regularly* provided that to each positive integer m there corresponds an integer n' such that for $n > n'$ the first m columns of the matrix $\| a_{is}(n) \|$ remain invariant. If a regular matrix $\| a_{is}(n) \|$ converges regularly to the limit $\| a_{is} \|$, the limit is obviously regular. In terms of regular convergence we now have the following useful criterion.

Let

$$y_i(n) = \sum_s a_{is}(n) x_s \tag{24}$$

be an infinite sequence of transformations, each determining a change of basis and let

$$x_i = \sum_s b_{is}(n) y_i(n)$$

be the corresponding sequence of inverse transformations. Then a sufficient condition that the limit of (24) *exist and determine a change of basis is that the two variable matrices* $\| a_{is}(n) \|$ *and* $\| b_{is}(n) \|$ *be regularly convergent.*

For under the conditions of the theorem the limiting matrices $\| a_{is} \|$ and $\| b_{is} \|$ will both be regular. Moreover, it is easy to see that the elements b_{is} of the second matrix will satisfy relations (13) and (14) of § 5. For suppose we wish to show that they satisfy any particular one of these relations, such as relation (13) corresponding to some particular choice of the indices i and j. Then, if the last non-vanishing element of the j-th column of the matrix $\| a_{ij} \|$ is the α'th we may choose n so large that the first α elements in the i-th row of $\| b_{is}(n) \|$ have reached their

final, limiting values and that all the elements in the j-th column of $\| a_{is}(n) \|$ have also reached their final values. We shall then have

$$\sum_s b_{is}\, a_{sj} = \sum_s b_{is}(n)\, a_{sj}(n) = \delta_{ij},$$

because $\| b_{is}(n) \|$ is, by hypothesis, the inverse of $\| a_{sj}(n) \|$. By a similar argument, the other relations in (13) and (14) may be shown to hold. Consequently, the regular matrix $\| b_{is} \|$ must be the inverse of the regular matrix $\| a_{is} \|$, and the transformation

$$y_i = \sum_s a_{is} x_s \tag{25}$$

must determine a change of basis, by the theorem of § 5.

It may be well to observe that the regular convergence of the matrix $\| a_{is}(n) \|$ is not, of itself, sufficient to insure that the limiting transformation (25) shall determine a change of basis. For example, the matrix of the transformation

$$\begin{aligned} y_1(n) &= x_1, \\ y_2(n) &= -x_{i-1} + x_i, && (1 < i < n), \\ y_i(n) &= x_i, && (i \geqq n), \end{aligned}$$

converges regularly to the matrix of the transformation represented by relation (16) in § 6. But we have seen that this last transformation does not determine a change of basis.

C. Sub-groups and factor groups.

8. A sub-group H of G will be said to be *closed* if its group space Σ_H is a closed sub-space of the group space Σ_G of G. If H is any sub-group whatever of G there is always a minimal closed sub-group H' of G containing the group H. The group space $\Sigma_{H'}$ of H' consists of all points and limit points of the group space Σ_H of H. Since the space Σ_G is metric, the space $\Sigma_{H'}$ is necessarily closed.

Let p_s, $(s = 1, 2, 3, \cdots)$, be any finite or infinite sequence of elements of G such that the norms of these elements are monotonically decreasing,

$$\omega(p_1) > \omega(p_2) > \omega(p_3) > \cdots.$$

Then the set of all linear combinations of the elements p_s is a closed sub-group of G.

Proof. The set of all linear combinations of the elements p_s is obviously a group H. Therefore, we have only to show that if a sequence of linear combinations

(26) $$q_i = \sum_s a_{is} p_s$$

approaches a limit q then q may always be represented as a linear combination of the elements p_s. We shall prove that for each value of s the coefficient a_{is} approaches a limiting value a_s; whence we shall at once be able to write

(27) $$q = \sum_s a_s p_s .$$

If we form the differences,

(28) $$q_{i+j} - q_i = \sum_s (a_{i+j,s} - a_{is}) p_s , \qquad (j \geqq 1),$$

the norm of each difference will be equal to the norm of the first non vanishing term in the corresponding right hand member of (28), *because the norms of the elements p_s are monotonically decreasing.* Therefore, if $q_{i+j} - q_i$ is to converge to zero, (as it must if q_i is to converge), each coefficient $a_{i+j,s} - a_{is}$ must ultimately become and remain zero, or in other words each coefficient a_{is} must attain a limiting value a_s. But, if each coefficient a_{is} attains a limiting value a_s, the difference $q - q_i$ obviously converges to zero; that is to say, q_i converges to q.

We have stressed the proof of this rather obvious theorem because it is so easy to jump to false conclusions when dealing with infinite groups of the sort here considered. Thus, for example, we cannot replace the condition that the norms $\omega(p_s)$ decrease monotonically by the weaker condition that they merely converge to zero. This is shown by the second example in § 6. If the elements y_i in relations (18) determined a closed group H the group H would have to be the group G itself. For, by (19), every finite linear combination of the x_i's is expressible in terms of the y_i's and every infinite linear combination is a limit of finite linear combinations. We have seen, however, that the group determined by the y_i's does not include the element $x_1 + x_3 + x_5 + \cdots$.

9. A finite or infinite sequence of elements p_i of G with monotonically decreasing norms will be called a *normal basis* of the closed sub-group of G consisting of all linear combinations of the p_i's.

Every closed sub-group H of G possesses a normal basis, one of which may always be found in the following manner. Corresponding to each value of n let us consider all elements of the group H of norm $1/n$. If there are any such elements at all, let p_n be one of them, so chosen that the numerical value of its leading coefficient (the nth) is as small as possible. We shall prove that the elements p_n arranged according to ascending values of n are a normal basis of H. Of course, our notation is such that p_n need not be defined for all values of n, but this is quite immaterial.

The proof consists in showing that the closed group H' consisting of all linear combinations of the elements p_n is identical with the group H. Now, every finite linear combination of the p_n's belongs to H, since H has the group property; therefore, also, every infinite linear combination belongs to H, since H is closed. Thus, the group H must contain the group H'. But the group H' must also contain the group H. For let q be any element of H. Then since the group H' is closed and since the distance function δ can only take on values of the form 0 or $1/n$ there must be an element p of H' such that $\delta(p, q)$ is a minimum. Now, if p were different from q we would have

$$\delta(q, p) = \omega(q-p) = \frac{1}{n}.$$

Therefore, the leading coefficient of the form $q-p$ would have to be exactly divisible by the leading coefficient of p_n, otherwise we could find an element of H of the form $q-p-\lambda p_n$ which would also be of norm $1/n$ but which would have a leading coefficient numerically smaller than that of p_n, contrary to the definition of p_n. But if the leading coefficient of $q-p$ were exactly divisible by the leading coefficient of p_n we could find a multiplier μ of p_n such that the element $q-p-\mu p_n$ would lack the terms in x_n and would therefore have a norm smaller than $1/n$. The element $p+\mu p_n$ of H' would consequently be closer to q than the element p, contrary to the definition of p. Thus, every element q of H must be identical with an element p of H'. This completes the argument.

If we re-define the distance between two points of the group space Σ_H in terms of the p_n's rather than in terms of the x_i's we obviously leave invariant the topological structure of the space Σ_H.

10. Every sub-group H of G determines a certain *factor group*, $G \pmod H$. Let us recall, very briefly how this factor group is defined. We arrange the elements of G into classes, called, *co-sets*, such that two elements belong to the same co-set if, and only if, their difference belongs to the sub-group H. These co-sets are the elements of the group $G \pmod H$. We shall write the symbol $[p]$ to denote the element of $G \pmod H$ containing the element p of G. There will thus be as many different symbols for the element $[p]$ of $G \pmod H$ as there are elements of G in the co-set containing p. The *negative* of the element $[p]$ is, by definition, the element $[-p]$; the *sum* of the elements $[p]$ and $[q]$ is the element $[p+q]$. It is easy to verify that the negative of an element and the sum of two elements are uniquely determined. The *zero element* of the group $G \pmod H$ is, of course, the co-set made up of all elements of H.

We shall confine our attention to the case where the sub-group H is closed. Each element $[p]$ of the factor group $G \pmod H$ will then be represented

by a *closed* point set in the group space Σ_G, for the point set in question will be the image of the space Σ_H under the displacement of the space Σ_G determined by the element p. We shall define the norm $\omega[p]$ of an element $[p]$ of the factor group as the greatest lower bound of the norm of the elements of G in the co-set $[p]$. Moreover, we shall define the *distance* between two elements of the factor group as the norm of their difference:

$$\delta[p][q] = \omega[p-q].$$

We shall prove in just a moment that the distance function $\delta[p][q]$ has properties analogous to the distance function $\delta(p, q)$ of § 4, and, hence, that the group space Σ_G is a metric space.

Let us remark, in passing, that if the sub-group H had not been assumed to be closed the norm of a non-vanishing element $[p]$ of the factor group might have been equal to zero.

D. Classification of factor groups.

11. Two factor groups F and F' will be said to be *homeomorphic* if, and only if, the following two conditions are fulfilled:

(i) The groups are simply isomorphic.

(ii) The isomorphism is of such a character that it induces a homeomorphic (i. e., a one-one bi-continuous) correspondence between the points of the two group spaces Σ_F and $\Sigma_{F'}$.

We shall now attack the main problem of the paper which is:

To characterize the factor groups of the group G by numerical invariants in such a way that two factor groups are homeomorphic if, and only if, they have the same invariants.

Let H be any closed sub-group of G and let

$$p_i = \sum_s a_{is} x_s \tag{29}$$

be any normal basis of H, (§ 6). Then the structure of the factor group G (mod H) is completely determined by the matrix $\| a_{is} \|$. We shall call the matrix $\| a_{is} \|$ a matrix *of* the factor group G (mod H). The matrices of two factor groups will be said to be *equivalent* if and only if the two groups are homeomorphic. To analyze the structure of the factor group G (mod H) we shall start with an arbitrary matrix of the group and try to find an equivalent matrix of as simple a form as possible.

Let us denote the leading coefficients of the forms p_i in (29) by α_i respectively. Then, as a first simplification, we may assume that all the

coefficients α_i are positive, since we are permitted to replace any of the forms p_i by their negatives.

Since the forms p_i are a normal basis of the sub-group H, no column of the matrix $\| a_{is} \|$ can contain more than one of the coefficients α_i. Moreover, the column containing the coefficient α_{i+1} must always follow the column containing the coefficient α_i. We shall call the columns containing the coefficients α_i the *principal columns* of the matrix $\| a_{is} \|$ as distinguished from the *secondary columns.* Moreover we shall call the marks x_s associated with the principal columns, *the principal marks* of the basis. Finally we shall call the coefficients α_i the *principal coefficients* of the matrix.

Now, in the notation of § 10, every element of the factor group $G \pmod H$ is represented by a symbol $[\sum c_s x_s]$, where the linear form appearing between the brackets is only determined to within a linear combination of the forms p_i. It will be convenient to reduce the symbol $[\sum c_s x_s]$ to a unique *canonical* type such that the coefficient c_s of each *principal* mark x_s is a non-negative integer smaller than the principal element α_i appearing in the column associated with x_s. This may evidently be accomplished by subtracting from the form $\sum c_s x_s$ a suitable linear combination of the forms p_i. If the symbol $[\sum c_s x_s]$ is in canonical form there is clearly no element of G of the co-set $[\sum c_s x_s]$ nearer to the origin than $\sum c_s x_s$; hence, by the definition of the norm of $[\sum c_s x_s]$, we must have

$$\omega [\sum_s c_s x_s] = \omega (\sum_s c_s x_s).$$

It follows, at once, that the inequality

$$\omega [p] > \omega [q]$$

implies the equality

$$\omega [p + \lambda q] = \omega [p], \qquad \lambda \neq 0,$$

which is the analogue of relation (2). Hence, by precisely the same argument as the one used in § 3 we find that *if* $[p]$, $[q]$, *and* $[r]$ *are any three points of the group space* $\Sigma_{G \pmod H}$ *then either all three distances* $\delta [p] [q]$, $\delta [p] [r]$ *and* $\delta [q] [r]$ *are equal or else two of them are equal to one another but greater than the third.*

The distance function δ therefore satisfies the triangle axiom; whence, *the group space* $\Sigma_{G \pmod H}$ *is a metric space.*

12. If one of the principal elements α_i has the value unity and if all elements of the matrix $\| a_{is} \|$ in the same row or column as α_i have the value zero except α_i itself we shall say that the row and column containing α_i are *redundant. If we strike out all redundant rows and columns from the matrix* $\| a_{is} \|$ *the resultant matrix will clearly be equivalent to* $\| a_{is} \|$.

For the canonical symbols $[\sum c_s x_s]$ representing the elements of the factor group $G(\text{mod } H)$ will obviously be identical with the symbols for the elements of the factor group determined by the new matrix. Moreover, the symbols will combine in the same way whether they represent elements of $G(\text{mod } H)$ or of the new factor group. Strictly speaking the norm of an element $[\sum c_s x_s]$ of the new group will, in general, be greater than the norm of the corresponding element of $G(\text{mod } H)$, owing to the elimination of the redundant columns of $\|a_{is}\|$ but this change in the norm will evidently have no effect on the definition of convergence.

By reversing the above process, we may transform a matrix $\|a_{is}\|$ *into an equivalent one by interpolating any number of redundant rows along with the corresponding redundant columns,* provided, of course, we never interpolate more than a finite number of rows between any two consecutive rows of $\|a_{is}\|$.

The lemma about redundant rows and columns will now be used in proving the following theorem:

The matrix $\|a_{is}\|$ *is equivalent to a matrix* $\|c_{is}\|$ *such that the principal element in each row of* $\|c_{is}\|$ *is a positive prime number or unity.*

Proof. Each principal element α_i of $\|a_{is}\|$ which is not a prime number or unity may be expressed as the product of $k+1$ prime factors

$$\alpha_i = \pi_0 \pi_1 \cdots \pi_k. \tag{30}$$

Wherever such a leading element occurs we shall enlarge the matrix $\|a_{is}\|$ by interpolating k redundant rows immediately after the row containing α_i and the corresponding k redundant columns immediately after the column containing α_i. We shall thereby obtain a new matrix $\|b_{is}\|$ which will be equivalent to $\|a_{is}\|$, by the lemma. We may think of the columns of $\|b_{is}\|$ as associated with the base elements y_s of a new group G' analogous to G and of the rows as associated with the base elements q_i of a sub-group H' of G'. The matrix $\|b_{is}\|$ will then be a matrix of the factor group G' $(\text{mod } H')$ homeomorphic with G $(\text{mod } H)$.

Let us now examine a $k+1$ rowed minor μ of the matrix $\|b_{is}\|$ composed of a block of terms in consecutive rows and columns, and such that the element in the upper left hand corner of μ is the principal element α_i in (30). Evidently, the minor μ will be a diagonal matrix with the diagonal elements $\alpha_i, 1, 1, \cdots, 1$. Therefore by the classical theory of finite matrices with integer elements, there will be a change of basis τ_i of G' and a change of basis σ_i of H', involving only the marks p_s and q_i associated with the columns and rows of the minor μ, which will transform the minor μ into one with the prime numbers $\pi_0, \pi_1, \pi_2, \cdots, \pi_k$

along the main diagonal, with the numbers 1 along the diagonal immediately above the main diagonal, and with zeros everywhere else. Thus, the effect of the two transformations τ_i and σ_i will be to replace the principal elements $\alpha_i, 1, 1, \cdots, 1$ by a set of prime principal elements $\pi_0, \pi_1, \cdots, \pi_k$.

Now, the resultant of the sequence of transformations τ_i $(i = 1, 2, 3, \cdots)$, determines a change of basis of the group G', since it is of the type (γ) discussed in § 6. Similarly the resultant of the transformations σ_i determines a change of basis of the sub-group H'. Moreover these two changes of basis reduce the matrix $\|a_{is}\|$ to the desired matrix $\|c_{is}\|$ in which all non-prime leading coefficients α_i have been eliminated. We have, therefore, proved the theorem.

13. Next, we shall prove that the *matrix* $\|a_{is}\|$ *is equivalent to a matrix* $\|b_{is}\|$ *such that all the elements in the secondary columns of* $\|b_{is}\|$ *are zero.* The transformation carrying $\|a_{is}\|$ into $\|b_{is}\|$ will be derived as the limit of a series of simple transformations, each of which will reduce to zero a new element in one of the secondary columns. To avoid unnecessary complications, we shall assume, as we may in view of the last theorem, that the principal element in each row of the matrix $\|a_{is}\|$ is a prime number or unity.

Let the first secondary column of $\|a_{is}\|$ containing a non-vanishing element be the sth column, and let the last non-vanishing element in this column be the element a_{is} in the ith row. We may then replace the element a_{is} by zero in the following manner. Let the principal element in the same row as a_{is} be the prime element π_i, and let the principal column containing π_i be the rth column, $(r < s)$. Then, by a suitable change of basis of the form

$$\begin{aligned} x'_r &= ax_r + bx_s, \\ x'_s &= cx_r + dx_s, \end{aligned} \qquad ad - bc = \pm 1 \tag{31}$$

involving only the principal mark x_r and the secondary mark x_s we may obviously replace the rth and sth columns of the matrix $\|a_{is}\|$ by linear combinations of the two of them in such a way that π_i is replaced by the highest common factor of π, and a_{is}, *and that* a_{is} *is replaced by zero.* By a succession of similar transformations, we may, therefore, reduce the elements in the secondary columns to zero, one after the other. The special case where the complete reduction may be effected in a finite number of steps requires no further comment. However, we must examine more closely the general case where the complete reduction necessitates a passage to the limit.

Let the resultant of the first n transformations of type (31) be expressed as a function of n.

$$x_i(n) = \sum_s \lambda_{is}(n) x_s, \tag{32}$$

let the inverse of (32) be

$$x_i = \sum_s \mu_{is}(n) x_s(n), \tag{33}$$

and let the matrix into which $\|a_{is}\|$ is transformed by (32) be $\|a_{is}(n)\|$. Then, clearly, the matrix $\|\mu_{is}(n)\|$ varies co-grediently with $\|a_{is}(n)\|$, as n increases, and the matrix $\|\lambda_{is}(n)\|$ varies contra-grediently with $\|a_{is}(n)\|$. We shall prove that all three of these matrices are regularly convergent. It will then follow, by § 7, that the variable transformation (32) converges to a limit and that the limiting transformation is a change of basis of G carrying the matrix $\|a_{is}\|$ into the desired matrix $\|b_{is}\|$.

Let us first examine transformation (31) a little more closely. If it so happens that the element a_{is} which we wish to replace by zero is exactly divisible by the principal element π_i, all that we shall have to do will be to add to the sth column of the matrix $\|a_{is}\|$ a suitable multiple λ of the rth column. The requisite transformation (31) will, therefore, reduce to the simplified form

$$\begin{aligned} x'_r &= x_r - \lambda x_s, \\ x'_s &= x_s. \end{aligned} \tag{34}$$

It will, thus, leave all the marks of the basis x_i invariant except the principal mark associated with the rth column of $\|a_{is}(n)\|$. If the element a_{is} is not exactly divisible by π_i we shall not be able to reduce (31) to the form (34). In this case, however, the elements π_i and a_{is} will be mutually prime, (since π_i is a prime number), consequently the transformation will reduce the principal element in the rth column from π_i to unity. Every subsequent transformation involving the principal mark $x_i(n)$ associated with the rth column, will therefore have to be of the simplified type (34).

The regular convergence of the matrix $\|a_{is}(n)\|$ now follows immediately. By what we have just shown, a variable mark $x_r(n)$ associated with a principal column of $\|a_{is}(n)\|$ can undergo at most one transformation of type (31) irreducible to type (34). Therefore, as n increases, each principal column of $\|a_{is}(n)\|$ will either remain invariant or change exactly once for some one value of n. Moreover, each secondary column of $\|a_{is}(n)\|$ will remain invariant from the moment when all its elements have been reduced to zero. The regular convergence of $\|a_{is}(n)\|$ is, therefore, assured and, with it, the regular convergence of $\|\mu_{is}(n)\|$, because $\|\mu_{is}(n)\|$ varies co-grediently with $\|a_{is}(n)\|$.

Finally, to show that the matrix $\|\lambda_{is}(n)\|$ is regularly convergent, let us re-examine relation (32) from which the matrix was derived. Initially, the variable marks $x_i(n)$ in (32) will have the values $x_i(0) = x_i$ and,

therefore, the norms $\omega(x_i) = 1/i$. Moreover, as may at once be verified by induction, the norm of each principal mark $x_r(n)$ will remain *invariant* and equal to $1/r$ as n increases, but the norm of each secondary mark $x_s(n)$ will increase to a new value $1/r$ each time the mark $x_s(n)$ enters into combination with a principal mark $x_r(n)$ as the result of a transformation of type (31) irreducible to type (34). Now, given any positive integer k, let us carry out the reduction of the matrix $\| a_{is} \|$ to such a point that, at the moment when we are about to start operating on a new secondary column, the first k columns of $\| a_{is}(n) \|$ have attained their ultimate, invariative form. Then, clearly, at each further step of the reduction we shall either be making a change of basis of type (31) such that the norms of $x_r(n)$ and $x_s(n)$ are both less than $1/k$ or a change of basis of type (34) such that the norm of $x(n)$ in less than $1/k$. But the only effect that these transformations can have on the right hand members of equations (32) will be to alter the coefficients $\lambda_{is}(n)$ of the variables x_i subsequent to the kth one x_k. Therefore, the transformation will leave the first k columns of the matrix $\| \lambda_{is}(n) \|$ invariant. This establishes the regular convergence of the matrix $\| \lambda_{is}(n) \|$ and thus completes the argument.

On the strength of the last theorem let us choose the matrix $\| a_{is} \|$ so that all the elements in its secondary columns are zero. Let us further express the group G as the direct sum of two groups A and B, the first consisting of all linear combinations of the primary marks x_i of G and the second consisting of all linear combinations of the secondary marks. Then *we may, at once, express the factor group* G (mod H) *as the direct sum of the groups* A (mod H) *and* B. The matrix of A (mod H) will be obtainable from the matrix $\| a_{is} \|$ by merely striking out all the secondary columns of $\| a_{is} \|$. It will have zeros below the main diagonal, prime numbers and 1's (but never zeros) along the main diagonal, and arbitrary numbers above. In view of the topological applications which we have in view, we shall call the group B the *Betti group* of G (mod H) and the group $T = A$ (mod H) the *torsion group* of G (mod H).

The torsion group $T = A$ (mod H) is a uniquely determined sub-group of the group G (mod H); that is to say, it is independent of the particular reduced matrix $\| a_{is} \|$ used in defining it. Every element a of T is characterized by the fact that we can find a multiple λa of a, $(\lambda \neq 0)$, which is as close as we please to zero. For let $\pi_1, \pi_2, \pi_3 \cdots$ be the elements down the main diagonal of the matrix of T. Then, the multiple $\pi_1 \pi_2 \cdots \pi_k a$ of a may evidently be expressed in terms of the marks associated with columns subsequent to the kth. Therefore, as k increases indefinitely, the norm of $(\pi_1 \pi_2 \cdots \pi_k a)$ approaches zero. On the other

hand, if b is an element of $G \pmod H$ but not of T then the symbol for b must involve at least one of the secondary marks x_s. Consequently, the norm of λb, $(\lambda \neq 0)$, can never be less than $1/s$.

The Betti group B is evidently not uniquely determined, for the form of the matrix $\| a_{is} \|$ is not altered if we add an arbitrary linear combination of the primary marks x_r to one of the secondary marks x_s. However, the topological structure of B does not depend on the particular way in which B is chosen, since B is obviously homeomorphic with $G \pmod H$, reduced again mod T.

The structure of the group B is completely determined when we know the number of marks in its basis. It is, therefore, characterized by a single invariant P, (its Betti number), which will either be a non-negative integer or $\aleph_0$.

The structure of the torsion group T requires further analysis.

14. To determine the structure of the torsion group T let us consider one of the matrices $\| c_{is} \|$ of T such that every diagonal element of $\| c_{is} \|$ is either a positive prime number or unity. *We shall first reduce the matrix* $\| c_{is} \|$ *to such a form that each element* c_{is} *above the main diagonal is zero unless the diagonal element* c_{ii} *in the same row as* c_{is} *and the diagonal element* c_{ss} *in the same column as* c_{is} *are both different from unity and from one another.*

Suppose that in the matrix $\| c_{is} \|$, as originally chosen, an element c_{is} above the main diagonal is different from zero. Then, unless the diagonal elements c_{ii} and c_{ss} are equal to the same prime number (greater than unity) there exist multipliers λ and μ such that we have

$$\lambda c_{ii} + \mu c_{ss} = 1,$$

and, hence, multipliers $\lambda_{is} = \lambda c_{is}$ and $\mu_{is} = \mu c_{is}$ such that we have

$$\lambda_{is} c_{ii} + \mu_{is} c_{ss} = c_{is}.$$

We may, therefore reduce the element c_{is} to zero by subtracting λ_{is} times the ith column from the sth column and μ_{is} times the sth row from the ith row. Let x_s and p_i be the base marks associated with the columns and rows of $\| c_{is} \|$ respectively. Then if we make the following changes of bases,

$$x_i' = x_i + \sum_s \lambda_{is} x_s, \qquad (s > i), \tag{35}$$

and

$$p_i' = p_i - \sum_s \mu_{is} p_s, \qquad (s > i), \tag{36}$$

which are of the form (β) described in § 6, we shall simultaneously reduce to zero all elements c_{is} such that their corresponding diagonal elements c_{ii} and c_{ss} are not equal to the same prime number. The sums in (35) and (36) are of course to be extended over all the multipliers associated with the various elements c_{is} which we are reducing to zero.

After the matrix $\|c_{is}\|$ has been simplified in the manner just indicated there will be a redundant row and column corresponding to each element 1 appearing along the main diagonal, because every other element in the same row and column as a diagonal element 1 will have to vanish. We shall therefore be allowed to strike out all the rows and columns containing a diagonal element 1.

Now, let us say that a base mark x_i or p_i is of *class* π (π prime) if the diagonal element of the simplified matrix is π. Then, clearly, each base mark p_i of class π is a linear combination of base marks x_s of the same class π, because all the elements in the ith row of the matrix $\|c_{is}\|$ are zero except, possibly, the ones in the columns corresponding to the base marks x_s of class π. *The group T is, therefore, the direct sum of a set of groups T_π, one corresponding to each prime number π and such that the matrix of each group T_π is the one obtained from the matrix of the group T by striking out all rows and columns except the ones associated with the base marks of class π.* Each group T_π will be determined by a matrix of the same general form as $\|c_{is}\|$ except that all the diagonal elements of the matrix of T_π will be equal to the same prime number π.

15. To complete the discussion we must now determine the structure of the groups T_π. Consider any group T_π, let its matrix be $\|e_{is}\|$, and let the marks associated with the rows and columns of $\|e_{is}\|$ be r_i and z_s respectively. Then, in the notation of § 10, the group elements of T_π will be represented by symbols of the form

$$\zeta = [\sum_s a_s z_s]. \tag{37}$$

We shall measure the norm of a group element ζ of T_π in the metric determined by the marks z_s. Thus, if the symbol (37) is in canonical form and if the leading term in the brackets is $a_n z_n$ the norm of ζ will be $1/n$ (cf. § 11). Each column of the matrix $\|e_{is}\|$ determines an element $\zeta_s = [z_s]$ of the group T_π. To simplify the notation we shall express a general element ζ of T_π as a linear combination of the elements ζ_i with non-negative coefficients smaller than π:

$$\zeta = \sum_s a_s \zeta_s. \tag{38}$$

When we operate with symbols of the form (38) the coefficients a_s are, of course, not reduced *modulo* π in the ordinary sense. They are reduced successively to non negative integers smaller than π with the help of relations $r_i = 0$ determined by the rows of the matrix $\|e_{is}\|$.

From the form of expressions (38) it follows, at once, that *the group* T_π *is compact.* For consider any infinite sequence of elements η_i of T_π,

$$\eta_i = \sum_s a_{is}\,\zeta_s, \qquad (i = 1, 2, \cdots). \tag{39}$$

Since the coefficients a_{is} can only take on the values $1, 2, \cdots, \pi - 1$ there will always be an infinite sub-sequence of (39) such that the first coefficient a_{i1} will have a constant value a_1 for every element η_i of the sub-sequence. Similarly, there will be an infinite sub-sequence of the previous sub-sequence such that the second coefficient a_{i2} will have a constant value a_2, and so on. The form

$$\eta = \sum_s a_s\,\zeta_s$$

determined by the successive coefficients $a_1, a_2, \cdots$ will clearly represent a limit element of the sequence (39), since there will be elements η_i differing by as little as we please from η. The group T_π must, therefore, be compact. By a similar argument it can be shown that the entire torsion group T is compact. The group G (mod H) is, however, not compact unless its Betti number vanishes.

16. Let us now begin with the very special case where the matrix of the group T_π has the numbers π along the main diagonal, the numbers -1 along the diagonal immediately above the main diagonal, and zeros everywhere else. From the rows of the matrix we have, at once, the relations

$$\zeta_{s+1} = \pi\,\zeta_s, \qquad (s = 1, 2, 3, \cdots),$$

whence, also, the relations

$$\zeta_{s+1} = \pi^s\,\zeta_1.$$

It therefore follows that if the matrix of T_π is a finite matrix with n columns the group T_π will be a cyclic group of order π^n generated by ζ_1. If the matrix of T_π is an infinite matrix, every group element of T will be of the form

$$\zeta = \sum_s (a_s\,\pi^s)\,\zeta_1 = a_0\,\zeta_1 + a_1\,\pi\,\zeta_1 + a_2\,\pi^2\,\zeta_1 + \cdots. \tag{40}$$

By a slight extension of the meaning of a generator we shall say that in the infinite case, the group T_π is the *π-adic group generated by* ζ_1. The

π-adic group generated by ζ_1 is, more precisely, the closure of the infinite group generated by ζ_1. If the element ζ in (40) is expressed in canonical form the coefficients a_i will all be non-negative integers smaller than π. The set of all elements ζ obviously has the cardinal number of the continuum.

We shall prove that, in the general case, *the group T_π is the direct sum of a countable number of π-adic groups and of cyclic groups of orders π^n*, $(n = 1, 2, \cdots)$.

17. We shall find it convenient to introduce π-adic operators of the form

$$\alpha = \sum_s a_s \pi^s \tag{41}$$

and to represent the elements of the π-adic group generated by ζ_1 by the abridged symbols $\alpha\zeta_1$. Two π-adic operators α and β are to be regarded as *equal* if, and only if, the elements $\alpha\zeta_1$ and $\beta\zeta_1$ of the π-adic group are identical. Every operator α may, therefore, be reduced to a unique canonical form,

$$\alpha = \sum_s b_s \pi^s, \tag{42}$$

such that the coefficients b_s are non-negative integers smaller than π. The coefficients b_s may evidently be determined, one after another, by successively solving the diophantine equations

$$\begin{aligned} b_0 &= a_0 + \lambda_0 \pi, \\ b_1 &= (a_1 + \lambda_0) + \lambda_1 \pi, \\ &\cdots\cdots\cdots \\ b_s &= (a_s + \lambda_{s-1}) + \lambda_s \pi, \\ &\cdots\cdots\cdots \end{aligned}$$

An integral operator may be regarded as a special case of a π-adic operator, but there will, of course, be π-adic operators which correspond to no integer, since we have seen that the π-adic operators have the cardinal number of the continuum.

If the leading coefficient of a π-adic operator α, written in the canonical form (42), is the nth one, b_n, we shall say that the norm of α is $1/n$. Thus, in particular, if the norm of an integral operator α is $1/n$ it means that the integer α is exactly divisible by π^{n-1}, but not by π^n.

The operators α may be combined by *addition*, *subtraction*, and *multiplication* in the way that ordinary power series are combined. Clearly the norm of a product $\alpha\beta$ can never exceed the norm of either factor. Therefore, it will never be possible to divide a π-adic operator α by

a π-adic operator β of smaller norm than α. However, *if the norm of β is equal to or greater than the norm of α there will always be a π-adic number β/α.* For it will always be possible to find a π-adic number γ_1 such that the leading term in the canonical expansion of $\gamma_1 \alpha$ is the same as the leading term in the canonical expansion of β. Hence it will always be possible to write

$$\beta = \gamma_1 \alpha + \beta_1$$

where the norm of the remainder β_1 is smaller than the norm of the dividend β. The quotient β/α will, therefore, be obtainable by the ordinary process of long division in the form of a convergent series,

$$\frac{\beta}{\alpha} = \gamma_1 + \gamma_2 + \gamma_3 + \cdots$$

of π-adic operators with monotonically decreasing norms.

18. We are now ready to analyze a perfectly general group of type T_π. Let

$$\zeta = a_n \zeta_n + \cdots, \qquad (0 < a_n < \pi),$$

be any non-vanishing element of T_π, expressed in canonical form. Then, clearly, the norm of the πth multiple $\pi\zeta$ of ζ is smaller than the norm of ζ itself. For if we express $\pi\zeta$ in terms of the elements ζ_s the coefficient of ζ_n will be exactly divisible by π and will, therefore, reduce to zero when we write $\pi\zeta$ in canonical form. It follows, at once, that every infinite series, of the type

$$\sum_s a_s \pi^s \zeta = a_0 \zeta + a_1 \pi \zeta + \cdots$$

converges to a definite element η of T_π. With the help of the π-adic operator (41) we shall be able to express η as $\alpha\zeta$.

A *π-adic basis* of the group T_π will be any set of elements η_s of T_π such that the following two conditions are fulfilled:

(i) Given any integer n there are, at most, a finite number of elements η^s of norms greater than $1/n$.

(ii) Every element ζ of T_π is expressible as a linear combination of the elements η_s with π-adic coefficients.

In view of condition (i) every linear combination of the elements η_s with π-adic coefficients will be convergent and will, therefore, have a meaning. In general, there will, of course, be relations of the form

$$\sum_s \alpha_s \zeta_s = 0$$

with π-adic coefficients connecting the marks of a π-adic basis.

A cyclic group of order π^n has a π-adic basis consisting of a single element η satisfying a relation $\pi^n \eta = 0$. A π-adic group has a π-adic basis consisting of a single element η satisfying no relation of the form $\alpha\eta = 0$, $(\alpha \neq 0)$.

Let us note, for future reference, the following obvious lemmas:

The relation

$$\sum_s \alpha_s \zeta_s = 0, \qquad (\alpha_i \ \pi\text{-adic}), \tag{43}$$

implies the relation

$$\beta \sum_s \alpha_s \zeta_s = 0, \tag{44}$$

where β is any π-adic number.

For relation (44) may be expressed as a convergent sum of integral multiples of relation (43) such that each term of the sum corresponds to a term in the expansion of β in powers of π. In general, relation (44) does not imply relation (43).

If the norm of a coefficient α_s in (43) *is* $1/n$, *the multiplier β may be so chosen that the coefficient of $\beta\alpha_s$ in* (44) *is the integer* π^{n-1}. For the coefficient α_s has the same norm as the integer π^{n-1}. Therefore, we may choose β as the quotient π^{n-1}/α.

19. To fix matters, we shall choose as a π-adic basis of the group T_π the set of all marks ζ_s associated with the columns of the matrix $\|e_{is}\|$ of T_π. Let this basis be denoted by Z_0. Then, if the basis Z_0 contains an element ζ_t which is expressible as a linear combination of the other base elements of Z_0 the element ζ_t is clearly *redundant* in the sense that if we discard it from the set Z_0 we shall still have left a π-adic basis of T_π. We shall prove that there is always a subset of the elements ζ_s forming a π-adic basis of T_π and containing no redundant element. This subset may readily be obtained as follows. If the basis Z_0 contains redundant elements we always discard the first one ζ_{t_1}, appearing in the sequence ζ_s, thereby obtaining a new basis Z_1. If Z_1 contains redundant elements, we discard the first one ζ_{t_2}, $(t_2 > t_1)$, thereby obtaining a new basis Z_2, and so on. If the discarding process comes to an end after a finite number of steps we obtain, at once, the desired basis Z_n. Moreover, if the process goes on forever, it is easy to show that the residual elements of Z_0 which are not discarded at any stage of the process form a basis Z_∞ of the desired sort. For let ζ be any element of the group T_π and let

$$\zeta = \sum_s a_{ns} \zeta_s \tag{45}$$

be any representation of ζ in terms of the base elements of Z_n. (If Z_n does not contain the mark ζ_s we shall, of course, write $a_{ns} = 0$).

Then, by the classical device for compact sets which we have already employed so often, there exists a sub-sequence of the sums

$$\sum_s a_{ns}\zeta_s, \qquad (n = 1, 2, 3, \ldots),$$

such that the first coefficients α_n, converge to a π-adic number α_1. There also exists a sub-sequence of this sub-sequence such that the second coefficients converge to α_2, and so on. In terms of the coefficients $\alpha_1, \alpha_2, \cdots$ we shall now be able to express ζ in the form

$$\zeta = \sum_s \alpha_s \zeta_s, \tag{46}$$

since there exist relations (45) differing by as little as we please from (46). Evidently, relation (46) expresses ζ in terms of the base marks of Z_∞ alone, for if ζ_s is not a mark of Z_∞ the coefficients a_{ns} will all be zero from a certain point on, therefore the coefficient α_s in (46) will also be zero. Moreover the basis Z_∞ cannot contain a redundant element. For a redundant element ζ_n of Z_∞ would necessarily be a redundant element of all the Z_n's. But if the element were redundant for Z_i, $(i = 0, 1, 2, \cdots, n-1)$, it would have to be the *first* redundant element of Z_{n-1}. Therefore, it would not be present in Z_n, and we would be led to a contradiction.

20. In view of what we have just proved, we may always start with a basis Z of T_π without redundant elements. Let the elements of the basis Z be ζ_s. Then if there is any relation

$$\sum_s \alpha_s \zeta_s = 0 \tag{47}$$

with π-adic coefficients connecting the elements ζ_s, the norms of the coefficients α_s will all have to be smaller than unity. For if the norm of a coefficient α_s were unity there would be a π-adic number of the form $1/\alpha_s$, (§ 18). Therefore, we could multiply (47) by $1/\alpha_s$ and obtain a relation in which the coefficient of ζ_s was unity. The element ζ_s would, therefore, be redundant, contrary to hypothesis.

A relation (47) will be said to be of *species* $1/n$ if no coefficient α_s has a norm greater than $1/n$ and if one coefficient, at least, has the norm $1/n$. By what we have just seen, the species of every relation (47) must be smaller than unity. In a relation of species $1/n$, the first element ζ_s with a coefficient α_s of norm $1/n$ will be called the *principal element* of ζ_s. If we multiply such a relation by $1/\alpha_s$ (cf. § 18), we shall evidently obtain an equivalent relation in which the coefficient of the principal element is the integer π^{n-1}.

Now suppose there is a relation of species $1/n_1$ among the base elements ζ_s, but no relation of greater species than $1/n_1$. Then, in the sequence ζ_s, let us pick the first element ζ_{t_1} which is a principal element of some relation of species $1/n_1$. The element ζ_{t_1} clearly appears in a relation of the form

$$\pi^{n_1-1}\zeta_{t_1}+\sum_{s\neq t_1}\alpha_s\zeta_s=0.$$

Moreover, since all the π-adic coefficients α_s are exactly divisible by π^{n_1-1}, we may write this relation as follows:

$$\pi^{n_1-1}\left(\zeta_{t_1}+\sum_{s\neq t_1}\beta_s\zeta_s\right)=0, \qquad \left(\beta_s=\frac{\alpha_s}{\pi^{n_1-1}}\right).$$

Thus if we make a change of basis which merely replaces ζ_{t_1} by

$$\zeta^1_{t_1}=\zeta_{t_1}+\sum_{s\neq t_1}\beta_s\zeta_s, \tag{48}$$

the new base element will evidently satisfy the relation

$$\pi^{n_1-1}\zeta^1_{t_1}=0.$$

Moreover, it will satisfy no relation in which the norm of its coefficient is greater than $1/n_1$. Otherwise the original base element ζ_{t_1} would have to satisfy a similar relation, as we could see, at once, by applying the inverse of transformations (48). The group T_π is, therefore, the direct sum of a cyclic group K_1 of order π^{n_1-1} generated by $\zeta^1_{t_1}$ and of a group T^1_π generated by the remaining base elements ζ_s, $(s\neq t_1)$.

In general, there will also be relations among the base elements of T^1_π, in which case, we shall decompose this new group into the direct sum of a cyclic group K_2 of order π^{n_2-1}, $(n_2\geqq n_1)$, and of a group T^2_π, and so on. If the process comes to an end after a finite number of steps we shall ultimately resolve the group T_π into the direct sum of a set of finite cyclic groups K_i and of a group T^n_π such that there are no relations whatever between the base elements of T^n_π. Moreover, the group T^n_π will obviously be the direct sum of a set of π-adic groups, each generated by one of the base elements of T^n_π.

21. There remains the case where the process of taking out finite cyclic groups goes on forever. To handle this case, a certain amount of care must be exercised. If the group T_π is to be the direct sum of the cyclic groups K_i and the residual group T^ω_π generated by the residual base elements, the norms of the base elements $\zeta^1_{t_i}$ of the groups K_i will have

to converge to zero, since the combined base elements of the groups K_i and T_π^ω will have to form a π-adic basis of the group T_π. But if the groups K_i are obtained according to the process indicated above there is no assurance that the norms of the base elements ζ'_{t_i} will converge to zero. The difficulty may readily be overcome in the following manner. If no two of the base elements ζ'_{t_i} have the same norm their norms will clearly converge to zero, and no difficulty will be encountered. If two or more elements ζ'_{t_i} do have the same norm, let K_v be the first cyclic groups such that the norm of its base element ζ'_{t_v} is the same as the norm of the base element ζ'_{t_u} of some group K_u already in the sequence. Then there always exists an integer k $(< \pi)$ such that the norm of $\zeta'_{t_v} - k\zeta'_{t_u}$ is smaller than the norm of ζ'_{t_v}, as may be seen, at once, by expressing ζ'_{t_u} and ζ'_{t_v} in canonical form and noting that the leading coefficient of $k\zeta'_{t_u}$ is k times the leading coefficient of ζ'_{t_u} reduced *modulo* π. What we do, therefore, is merely to replace the cyclic group K_v by a new group K'_v generated by $\zeta'_{t_v} - k\zeta'_{t_u}$. The direct sum of K_u and K_v is clearly identical with the direct sum of K_u and K'_v. For the matrix of $K_u + K_v$ is of the form

$$\begin{vmatrix} \pi^\alpha & 0 \\ 0 & \pi^{\alpha+\beta} \end{vmatrix}, \qquad (\beta \geqq 0). \tag{49}$$

Moreover, when we replace ζ'_{t_v} by $\zeta'_{t_v} - k\zeta'_{t_u}$ the matrix becomes

$$\begin{vmatrix} \pi^\alpha & 0 \\ k\pi^{\alpha+\beta} & \pi^{\alpha+\beta} \end{vmatrix}$$

which reduces, at once, to the original form (49) if we subtract from the relation corresponding to the last row the $k\pi^\beta$th multiple of the relation corresponding to the first row. After a finite number of steps similar to the one just made, we may obviously replace the group K_v by a group K_v^* such that the norm of its generator is different from the norms of the generators of all the preceding groups K_i. In other words, there is no loss of generality in assuming that the groups K_i are successively chosen in such a manner that no two of their generators have the same norm.

Now, when the group elements K_i have been chosen in the manner last indicated it is easy to see that the group T_π is the direct sum of the groups K_i and the group T_π^ω generated by the residual base elements. The proof follows the lines of previous proofs so closely that a rapid outline will be sufficient. Let ζ be any element of the group T_π and let

$$\zeta = \sum_{i=1}^{n} \alpha_{ni}\zeta'_{t_i} + \sum_{s=1}^{\infty} \beta_{ns}\zeta_s \tag{50}$$

be any expression for ζ in terms of the base elements of the groups $K_i (i \leqq n)$ and the group T_π^n. (Of course, the coefficient β_{ns} vanishes if ζ_s is not a base element of T_π^n). Then, there is always a sub-sequence of (50) which converges to a form

$$\zeta = \sum_{i=1}^{\infty} \alpha_i \zeta_{t_i}' + \sum_{s=1}^{\infty} \beta_s \zeta_s, \tag{51}$$

where β_s is zero if ζ_s is not a base element of T_π^ω. This proves that the group T_π is the sum of the groups K_i and the group T_π^ω. To prove that T_π is the *direct sum* we observe that a relation of the form (51) in which one of the coefficients a_i was smaller than the period of K_i would imply a similar relation (50), which would contradict the fact that T_π was the direct sum of K_i $(i \leqq n)$ and T_π^n.

22. The argument is now nearly completed. It may, of course, happen that there are relations among the base elements of the residual group T_π^ω, in which case, we can start afresh and remove cyclic groups from T_π^ω. Let K_ω be the first cyclic group to be removed. Then, clearly, the order of K_ω must be *greater* than the orders of all the groups K_i previously removed. For the presence in T_π^ω of a cyclic group of the same order as a K_i would imply that at some stage in the formation of the groups K_i we had chosen one of the groups K_i incorrectly by failing to pick the *first* principal element ζ_{t_i} occurring in a relation (47) of maximal species. We may, therefore, choose K_ω in such a way that the norm of its base element is different from the norms of the base elements of all the groups K_i, and so on. Thus, by repeating the entire process an infinite number of times, if necessary, we must ultimately arrive at a countable set of cyclic groups

$$K_1, K_2, \cdots, K_\omega, K_{\omega+1}, \cdots,$$

such that the norms of their base elements are all different, and at a group T_π^* determined by whatever base elements ζ_s of T_π are still left after all possible cyclic groups of all possible orders have been taken out. By a mere repetition of the argument made above, the group T_π may be shown to be the direct sum of the cyclic groups K_i and the group T_π^*. Finally, the group T_π^* is obviously the direct sum of a set of π-adic groups, one determined by each of its base elements. The resolution of the group T_π is, therefore, completed.

23. In conclusion, we evidently need the following invariants to characterize the factor group $G \pmod H$ completely.

(i) A Betti number P.

(ii) An invariant $[\pi^\infty]$, for every choice of the positive prime number π, measuring the number of π-adic groups in the decomposition of T.

(iii) An invariant $[\pi^n]$, for every choice of the positive integer n and the prime π, measuring the number of cyclic groups of order π^n in the decomposition of T.

The numbers P, $[\pi^\infty]$, and $[\pi^n]$ are either non negative integers or infinite. We have already proved that the number P is an invariant, § 13. The numbers $[\pi^\infty]$ and $[\pi^n]$ may be given the following invariantive significance. Let F_π be the sub-group of T_π consisting of all elements of finite period, and let F_π^n be the sub-group of T_π consisting of all elements of period π^n or less. Then the number of independent elements in the factor group $T_\pi \pmod{F_n}$ is clearly $[\pi^\infty]$, and the number of elements in the factor group $F_\pi^n \pmod{F_\pi^{n-1}}$ is $n^{[\pi^n]}$.

If two groups $G \pmod{H}$ and $G \pmod{H'}$ have the same invariants they are obviously homeomorphic.

ÜBER EINEN SATZ VON HERRN M. H. STONE[1].

VON J. v. NEUMANN, PRINCETON.

1. Herr M. H. Stone hat vor einiger Zeit den folgenden Satz formuliert und bewiesen[2]:

(*S*) *Sei U_t eine Schar unitärer Operatoren des Hilbertschen Raumes, mit der Gruppeneigenschaft*

$$U_t U_s = U_{t+s}.$$

(*t, s reelle Zahlen*). *Dann existiert eine Zerlegung der Einheit $E(\lambda)$[3], so daß symbolisch*

$$U_t = \int_{-\infty}^{+\infty} e^{it\lambda}\, d\, E(\lambda)$$

gilt[4].

Hierbei ist Voraussetzung, daß U_t stetig von t abhängt[5]. Die Anwendungen, die von diesem Satze, besonders in der von Herrn B. O. Koopman entdeckten Operatorenbehandlungsweise der klassischen Mechanik gemacht werden können[6] (so konnte der Verfasser mit ihrer Hilfe die klassisch-mechanische Quasi-Ergodenhypothese beweisen[7]), lassen es als erwünschenswert erscheinen, den Stoneschen Satz auch ohne die Stetigkeitsannahme zu beweisen[8]. Auch rein mathematisch ist eine solche Verschärfung anzustreben.

In einer früheren Abhandlung[9] führte der Verfasser den Begriff der meßbaren Abhängigkeit eines Operators A_t vom Zahlenparameter t ein:

---

[1] Received March 16, 1932.

[2] Proc. Nat. Ac. 16, February 1930, S. 173–174.

[3] Vgl. z. B. die Abhandlung des Verfassers „Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren", Math. Ann. 102,1 (1929). Sie soll im folgenden als E zitiert werden.

[4] D. h. für zwei beliebige f, g des Hilbertschen Raumes

$$(U_t f, g) = \int_{-\infty}^{+\infty} e^{it\lambda}\, d\, (E(\lambda) f, g).$$

Die Bezeichnung nach a. O. Anmerkung [2], sowie E.

[5] Vgl. z. B. die Abhandlung des Verfassers „Zur Algebra der Funktionaloperatoren usw.", Math. Ann. 102,3 (1929), wo der Stetigkeitsbegriff variabler Operatoren präzisiert wird, insbesondere S. 381–388. Diese Arbeit soll im folgenden als A zitiert werden.

[6] Proc. Nat. Ac. 17, May 1931.

[7] Proc. Nat. Ac. 18, January 1932, sowie weitere Abhandlungen in den Proceedings, und den zweitfolgenden Artikel dieses Annals-Heftes.

[8] Vgl. den in Anmerkung [7] genannten Artikel des Verfassers in diesem Annals-Bande.

[9] Math. Ann. 104,4 (1931), S. 572.

(*M*) *Der Operator A_t hängt meßbar vom Zahlenparameter t ab, wenn für alle f, g des Hilbertschen Raumes die numerische Funktion $(A_t f, g)$ im gewöhnlichen (Lebesgueschen) Sinne meßbar ist. Übrigens kann man bei dieser Definition f, g auf die Elemente eines vollständigen normierten Orthogonalsystems $\varphi_1, \varphi_2, \cdots$ beschränken*[10].

Im nächsten Paragraphen zeigen wir, daß die Prämissen von (*S*) (d. i.: U_t unitär, $U_t U_s = U_{t+s}$) zusammen mit (*M*) (d. i.: U_t meßbar von t abhängig) die Stetigkeit von U_t in bezug auf t nach sich zieht. Im letzten Paragraphen geben wir einen neuen Beweis des Stoneschen Satzes an.

2. Erfülle U_t die Prämisse von (*S*) sowie (*M*). Da $(U_t f, g)$ meßbar von t abhängt, (f, g beliebig, aber vorläufig fest im Hilbertschen Raume), gibt es nach einem Satze von Lusin[11] zu jedem $\varepsilon > 0$ eine t-Menge $\mathfrak{T} = \mathfrak{T}_{f,g,\varepsilon}$ von Lebesgueschem Maße $< \varepsilon$, so daß $(U_t f, g)$, als Funktion des auf die Komplementärmenge von $\mathfrak{T}$ beschränkten t betrachtet, stetig ist. Sei $f_1, f_2, \cdots$ eine im Hilbertschen Raume überall dichte Folge[12], durchlaufe f, g alle Paare f_m, f_n. Wir setzen dazu $\varepsilon = \frac{\delta}{2^{m+n}}$ $(\delta > 0)$. Die Menge $\mathfrak{T}' = \mathfrak{T}'_\delta = \sum_1^{\infty m,n} \mathfrak{T}_{f_m, f_n, \frac{\delta}{2^{m+n}}}$ hat ein Maß $< \sum_1^{\infty m,n} \frac{\delta}{2^{m+n}} = \delta$, und für das auf ihre Komplementärmenge beschränkte t sind alle $(U_t f_m, f_n)$ stetig, also auch alle $(U_t f, g)$ (denn $(U_t f, g)$ ist, wenn t als Parameter angesehen wird, gleichmäßig stetig in f, g für alle t[13]). Wenn also t und $t_1, t_2, \cdots$ nicht in $\mathfrak{T}'$ liegen, und $t_n \to t$, so gilt für alle f, g $(U_{t_n} f, g) \to (U_t f, g)$, d. h. die $U_{t_n} f$ konvergieren schwach[14] gegen $U_t f$. Da aber gleichzeitig $\| U_{t_n} f \|$ gegen $\| U_t f \|$ konvergiert (wegen der Unitarität sind ja alle $= \| f \|$), konvergiert $U_{t_n} f$ sogar stark gegen $U_t f$[15]. Nach einem bekannten Satze von Lebesgue hat in allen Punkten t außerhalb von $\mathfrak{T}'$, bis auf eine Menge vom Maße 0, dieses $\mathfrak{T}'$ die Dichte 0[16]. Für ein solches t können

[10] D. h.: die Matrixelemente der A_t (im Koordinatensystem der $\varphi_1, \varphi_2, \cdots$) sollen meßbare Funktionen von t sein. Die Gleichwertigkeit der Definitionen beweist man leicht, vgl. a. a. O. Anmerkung [9].

[11] C. R., 154, 1912, S. 1688, oder Sierpinski, Fund. Math. 3, 1922, S. 320.

[12] Der Hilbertsche Raum ist separabel! Vgl. z. B. E, S. 65 und 109–111.

[13] Da es in f, g linear ist, genügt es, die Fälle $f = 0$ oder $g = 0$ zu betrachten. Dann ist $(U_t f, g) = 0$, und wegen der Schwarzschen Ungleichheit sowie der Unitarität von U_t

$$|(U_t f, g)| \leqq \| U_t f \| \cdot \| g \| = \| f \| \cdot \| g \|.$$

Vgl. E, S. 64, Satz 1.

[14] Vgl. z. B. A, S. 378–381.

[15] Wenn f_n schwach gegen f konvergiert, und $\| f_n \|$ (numerisch) gegen $\| f \|$, so gilt sogar starke Konvergenz:

$$\| f_n - f \|^2 = \| f_n \|^2 + \| f \|^2 - 2\,\Re\,(f, f_n) \to 2\,\| f \|^2 - 2\,\Re\,(f, f) = 0.$$

[16] Vgl. z. B. Carathéodory, „Reelle Funktionen", Berlin 1918, S. 492—497, Satz 3. Die dort auftretende Funktion $f(P)$ sei $f(t)$, u. zw. $= 1$ in $\mathfrak{T}'$, sonst $= 0$.

wir also zu jedem f, $\theta>0$, $\eta>0$ folgendes finden: Erstens ein $\delta_1>0$, so daß für $\delta' \leqq \delta_1$, >0 der im Intervalle $t-\delta'<t'<t+\delta'$ liegende Teil von $\mathfrak{T}'$ ein Maß $<\theta\cdot\delta'$ hat (wegen der Dichte 0!). Zweitens ein $\delta_2>0$, so daß für jedes nicht zu $\mathfrak{T}'$ gehörige t' des Intervalles $t-\delta_2<t'<t+\delta_2$ $\|U_{t'}f-U_tf\|<\eta$ gilt (wegen der Stetigkeit). Sei $\delta_0 = \text{Min}(\delta_1, \delta_2)>0$, also $\delta_0=\delta_0(t,f,\theta,\eta)$. Für $\delta' \leqq \delta_0$, >0 gilt also im Intervalle $t-\delta'<t'<t+\delta'$ überall, bis auf eine Menge vom Maße $<\theta\cdot\delta'$, $\|U_{t'}f-U_tf\|<\eta$. In dieser letzteren Aussage kommen $\mathfrak{T}'=\mathfrak{T}'_\delta$, δ nicht mehr vor, sondern nur U_t und t, f, θ, η sowie δ_0; wir wollen sie, die analoge Denjoysche Begriffsbildung verallgemeinernd, approximative starke Stetigkeit von U_t in t nennen (vgl. a. a. O. Anm. [11]). Sie gilt für alle t außerhalb von $\mathfrak{T}'$, bis auf eine Menge vom Maße 0, d. h. für alle t bis auf eine Menge vom Maße $<\delta$. Da aber $\delta>0$ beliebig war, für alle t, bis auf eine Menge vom Maße 0.

Wegen $U_t = U_{t-t_0}U_{t_0}$, $U_{t'} = U_{t-t_0}U_{t_0+(t'-t)}$ und der Unitarität von U_{t-t_0} ist U_t überall approximativ stark stetig, wenn es dies für t_0 ist — also ist ersteres der Fall.

Sei nun ein f und $\varepsilon>0$ gegeben, wir wählen ein $\delta>0$ so, daß für alle t' in $-\delta<t'<\delta$, bis auf eine Menge vom Maße $<\frac{1}{2}\delta$, $\|U_{t'}f-f\|<\frac{\varepsilon}{2}$ ist $\left(t=0,\ \theta=\frac{1}{2},\ \eta=\frac{\varepsilon}{2}\right)$. Ferner seien t_1, t_2 beliebig, aber mit $|t_1-t_2|<\delta$. Für $\left|\tau-\frac{t_1+t_2}{2}\right|<\frac{\delta}{2}$ ist bestimmt $|t_1-\tau|<\delta$, $|\tau-t_2|<\delta$, also gilt auch hier bis auf Mengen von Maßen $<\frac{\delta}{2}$, $\|U_{t_1-\tau}f-f\|<\frac{\varepsilon}{2}$, $\|U_{\tau-t_2}f-f\|<\frac{\varepsilon}{2}$. Bis auf eine Menge vom Maße $<\delta$ gilt daher beides zugleich, und da das Intervall $\left|\tau-\frac{t_1+t_2}{2}\right|<\frac{\delta}{2}$ das Maß δ hat, gibt es ein τ mit allen diesen Eigenschaften. Hieraus folgt aber:

$$\begin{aligned}\|U_{t_1}f-U_{t_2}f\| &\leqq \|U_{t_1}f-U_\tau f\|+\|U_\tau f-U_{t_2}f\|\\ &= \|U_\tau(U_{t_1-\tau}f-f)\|+\|U_{t_2}(U_{\tau-t_2}f-f)\|\\ &= \|U_{t_1-\tau}f-f\|+\|U_{\tau-t_2}f-f\|<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.\end{aligned}$$

Damit ist die starke Stetigkeit von U_t für alle t (vgl. a. a. O. Anm. [5]) bewiesen.

3. Wir gehen nun daran, den Stoneschen Satz zu beweisen.

Sei $a(t)$ eine numerische Funktion mit endlichem $\int_{-\infty}^{+\infty}|a(t)|\,dt$, dann

können wir einen Operator A symbolisch durch $\int_{-\infty}^{+\infty} a(t)\, U_t\, dt$ definieren, d. h. durch

$$(Af, g) = \int_{-\infty}^{+\infty} a(t)\ (U_t f, g)\ dt$$

(vgl. a. a. O. Anm. [9], S. 573–574). Man beweist leicht, wenn auch $\int_{-\infty}^{+\infty} |b(t)|\, dt$ endlich und $B = \int_{-\infty}^{+\infty} b(t)\, U_t\, dt$ ist, die folgenden Formeln:

$$A^* = \int_{-\infty}^{+\infty} \overline{a(-t)}\, U_t\, dt,\ cA = \int_{-\infty}^{+\infty} c\, a(t)\, U_t\, dt,$$
$$A+B = \int_{-\infty}^{+\infty} (a(t)+b(t))\, U_t\, dt,$$
$$U_s A = A U_s = \int_{-\infty}^{+\infty} a(t-s)\, U_t\, dt,$$
$$AB = \int_{-\infty}^{+\infty} c(t)\, U_t dt, \quad \left(c(t) = \int_{-\infty}^{+\infty} a(s)\, b(t-s)\, ds\right)^{17}$$

(vgl. die analogen Beweise am oben a. O.). Die letzteren Formeln zeigen, daß unsere $A, B, \cdots$ miteinander und mit den U_t vertauschbar sind.

Wir setzen nun $a(t) = \begin{cases} e^{-t} & \text{für } t \geqq 0 \\ 0 & \text{für } t < 0 \end{cases}$, d. h.

$$A = \int_0^{+\infty} e^{-t}\, U_t\, dt,\ A^* = \int_{-\infty}^{0} e^{t}\, U_t\, dt.$$

Daraus folgt sofort:

$$\left.\begin{aligned} A+A^* &= \int_{-\infty}^{+\infty} e^{-|t|}\, U_t\, dt, \\ AA^* = A^*A &= \int_{-\infty}^{+\infty} \tfrac{1}{2} e^{-|t|}\, U_t\, dt, \end{aligned}\right\} AA^* = A^*A = \tfrac{1}{2}(A+A^*).$$

Für $V = 1-2A$ ergibt das: $VV^* = V^*V = 1$, d. h. V ist unitär. Es ist, ebenso wie A, mit den V_t vertauschbar.

Betrachten wir den Ausdruck $\frac{U_t-1}{t}(V-1)$. Er ist gleich

$$-\frac{2}{t}(U_t-1)A = \frac{2}{t}(A - U_t A),$$

also gleich

$$\frac{2}{t}\left(\int_0^{\infty} e^{-s}\, U_s\, ds - \int_t^{\infty} e^{-s+t}\, U_s\, ds\right)$$
$$= 2e^t \cdot \frac{1}{t}\int_0^t e^{-s}\, U_s\, ds - 2\,\frac{e^t-1}{t}\cdot\int_0^{\infty} e^{-s}\, U_s\, ds.$$

[17] A^* ist die Adjungierte von A. Man beachte, daß aus $U_t U_s = U_{t+s}$ folgt. $U_0 = 1$, $U_t^{-1} = U_{-t}$, als wegen der Unitarität $U_t^* = U_t^{-1} = U_{-t}$.

Im ersten Glied konvergiert der erste Faktor für $t \to 0$ gegen 2, der zweite (der Operator!) wegen der Stetigkeit von U_s gegen $U_0 = 1$ (stark!), im zweiten Glied konvergiert der erste Faktor gegen 2, der zweite (der Operator!) ist gleich A. Also ist der (starke Operatoren-)Limes gleich $2 \cdot 1 - 2 \cdot A = V + 1$. Etwas anders geschrieben:

$$\frac{U_t - 1}{it}(V-1) \underset{\text{stark}}{\longrightarrow} -i(V+1) \qquad (t \to 0).$$

Aus dieser Formel folgt Verschiedenes. Erstens hat $Vf = f$ $(V-1)f = 0$ zur Folge, also

$$\frac{U_t - 1}{it}(V-1)f = 0, \; i(V+1)f = 0, \text{ d. h. } Vf = -f.$$

Also $f = 0$. Somit ist V die Cayleysche Transformierte eines hypermaximalen Operators R.[18] Rg hat Sinn, wenn $g = (V-1)f$ ist, und dann ist es gleich $-i(V+1)f$,[18] also gilt für alle g, für die Rg Sinn hat,

$$\frac{U_t - 1}{it} g \underset{\text{stark}}{\to} Rg \qquad (t \to 0).$$

Wenn $\frac{U_t - 1}{it} g$ auch nur einen schwachen Limes für $t \to 0$ hat, können wir diesen gleich $R'g$ setzen, dadurch einen Operator R' definierend. Aus

$$\left(\frac{U_t - 1}{it} g, h\right) = \left(g, \frac{U_t^* - 1}{-it} h\right) = \left(g, \frac{U_{-t} - 1}{i(-t)} h\right)$$

folgt $(R'g, h) = (g, R'h)$, so daß R' Hermitesch ist. Da es nach Obigem das hypermaximale R fortsetzt, ist $R' = R$. D. h.: wenn Rg keinen Sinn hat, konvergiert $\frac{U_t - 1}{it} g$ nicht einmal schwach.

Sei schließlich $E(\lambda)$ die zu R gehörige Zerlegung der Einheit.[19] Wir definieren eine Operatorenschar U_t' symbolisch durch $\int_{-\infty}^{+\infty} e^{it\lambda}\, dE(\lambda)$, d. h. durch

$$(U_t' f, g) = \int_{-\infty}^{+\infty} e^{it\lambda}\, d(E(\lambda)f, g),$$

ganz analog wie in einer früheren Abhandlung des Verfassers.[20] Durch dieselben Rechnungen wie dort[21] erkennt man:

[18] E, S. 80–81 und 91.
[19] E, S. 91, Def. 17 und S. 92, Satz 36.
[20] Annals of Math., 32/2 (1931).
[21] Vgl. insbesondere S. 205, a)—b), und deren Beweise.

$$U_0' = \int_{-\infty}^{+\infty} dE(\lambda) = 1,$$

$$U_t'^* = \int_{-\infty}^{+\infty} \overline{e^{it\lambda}}\, dE(\lambda) = \int_{-\infty}^{+\infty} e^{-it\lambda}\, dE(\lambda) = U'_{-t},$$

$$U_t' U_s' = \int_{-\infty}^{+\infty} e^{it\lambda} \cdot e^{is\lambda}\, dE(\lambda) = \int_{-\infty}^{+\infty} e^{i(t+s)\lambda}\, dE(\lambda) = U'_{t+s}.$$

Also ist $U_t' U_t'^* = U_t' U'_{-t} = U_0' = 1$, ebenso $U_t'^* U_t' = 1$, d. h. U_t' unitär, mit derselben Gruppeneigenschaft wie die U_t. Ferner ist jedes U_t mit V vertauschbar, also mit allen $E(\lambda)$,[22] also auch mit allen U_s'.

Wenn für $\lambda \geqq C$ $E(\lambda) f = f$ ist, und für $\lambda \leqq -C$ $E(\lambda) f = 0$ (was z. B. für jedes $f = (E(C) - E(-C))g$ der Fall ist[23]), so können wir bei $U_t' f$ das $\int_{-\infty}^{+\infty}$ durch ein $\int_{-C}^{+C}$ ersetzen. Dann erkennt man mühelos die Legitimität der folgenden Rechnung (das $\underset{\text{stark}}{\rightarrow}$ gilt für $t \rightarrow 0$):

$$\frac{U_t' - 1}{it} f = \frac{\int_{-C}^{+C} e^{it\lambda}\, dE(\lambda) f - \int_{-C}^{+C} dE(\lambda) f}{it}$$

$$= \int_{-C}^{+C} \frac{e^{it\lambda} - 1}{it}\, dE(\lambda) f \underset{\text{stark}}{\rightarrow} \int_{-C}^{+C} \lambda\, dE(\lambda) f$$

$$= \int_{-\infty}^{+\infty} \lambda\, dE(\lambda) f = Rf.$$

Für diese f hat Rf Sinn, und sie liegen überall dicht, da für $C \rightarrow +\infty$ $(E(C) - E(-C))\, g \rightarrow g - 0 = g$.[22] Somit gilt in einer überall dichten Menge der f: $\frac{U_t' - 1}{it} f \rightarrow Rf$, $\frac{U_t - 1}{it} f \rightarrow Rf$, also

$$i\left(\frac{U_t - 1}{it} f - \frac{U_t' - 1}{it} f\right) = \frac{U_t - U_t'}{t} f \rightarrow 0.$$

Für ein solches f ist aber:

$$\|(U_t - U_t')f\| = \left\| \sum_1^n \nu \left(U_{\frac{m+1-\nu}{n} t} U'_{\frac{\nu-1}{n} t} - U_{\frac{n-\nu}{n} t} U'_{\frac{\nu}{n} t} \right) f \right\|$$

$$= \left\| \sum_1^n \nu\, U_{\frac{n-\nu}{n} t} U'_{\frac{\nu-1}{n} t} \left(U_{\frac{t}{n}} - U'_{\frac{t}{n}} \right) f \right\|$$

$$\leqq \sum_1^n \nu \left\| U_{\frac{n-\nu}{n} t} U'_{\frac{\nu-1}{n} t} \left(U_{\frac{t}{n}} - U'_{\frac{t}{n}} \right) f \right\|$$

[22] Vgl. z. B. E, S. 115, Satz 10* und dessen Beweis, woraus dies leicht folgt.
[23] Vgl. die Formeln in E, S. 91, Def. 17.

$$= \sum_1^n \nu \left\| \left(U_{\frac{t}{n}} - U'_{\frac{t}{n}} \right) f \right\|$$
$$= n \left\| \left(U_{\frac{t}{n}} - U'_{\frac{t}{n}} \right) f \right\|$$
$$= t \cdot \left\| \frac{U_{\frac{t}{n}} - U'_{\frac{t}{n}}}{\frac{t}{n}} f \right\|.$$

Für $n \to +\infty$ strebt der zweite Faktor gegen 0, also ist

$$\|(U_t - U'_t) f\| = 0, \qquad U_t f = U'_t f.$$

Da $U_t f$, $U'_t f$ stetig von f abhängen, gilt dies nicht nur in der genannten überall dichten f-Menge, sondern für alle f. Also ist

$$U_t = U'_t = \int_{-\infty}^{+\infty} e^{it\lambda} \, dE(\lambda).$$

Damit ist die Behauptung von (S) bewiesen.

EINIGE SÄTZE ÜBER MESSBARE ABBILDUNGEN[1].

VON J. v. NEUMANN, PRINCETON.

1. Den Gegenstand dieser Arbeit bilden einige Sätze über meßbare Funktionen, die teilweise wegen ihrer Anwendung im unmittelbar nachfolgenden Artikel des Verfassers abgeleitet wurden. Immerhin sind sie vielleicht auch von selbständigem mathematischen Interesse.

Wir werden im folgenden gewisse Mengen $\mathfrak{Q}$ betrachten, die kurz m-Räume genannt werden sollen; sie sind so definiert:

DEFINITION 1. *Die Menge $\mathfrak{Q}$ heißt ein m-Raum, wenn sie metrisch, separabel und vollständig ist*[2] *(ihre Elemente sollen $x, y, \cdots$ heißen, die Entfernung $\overline{x,y}$).*

Beispiele von m-Räumen sind zahlreich: Der k-dimensionale Euklidische Raum, in dem die Distanz zweier Punkte $\{\xi_1, \cdots, \xi_k\}$, $\{\eta_1, \cdots, \eta_k\}$ (die ξ, η sind reelle Zahlen!) durch $\sqrt{(\xi_1-\eta_1)^2+\cdots+(\xi_k-\eta_k)^2}$ definiert wird, oder auch durch $\operatorname{Max}[|\xi_1-\eta_1|, \cdots, |\xi_k-\eta_k|]$; der Hilbertsche Raum (Punkte: $\{\xi_1, \xi_2, \cdots\}$, $\xi_1^2+\xi_2^2+\cdots$ endlich), in dem die Distanz zweier Punkte $\{\xi_1, \xi_2, \cdots\}$, $\{\eta_1, \eta_2, \cdots\}$ durch $\sqrt{(\xi_1-\eta_1)^2+(\xi_2-\eta_2)^2+\cdots}$ definiert wird; der ∞-dimensionale Raum[3] (Punkte: alle $\{\xi_1, \xi_2, \cdots\}$), in dem die Distanz zweier Punkte $\{\xi_1, \xi_2, \cdots\}$, $\{\eta_1, \eta_2, \cdots\}$ durch $\operatorname{Min}_{n=1,2,\cdots}\left\{\operatorname{Max}\left[|\xi_1-\eta_1|, \cdots, |\xi_n-\eta_n|, \frac{1}{n}\right]\right\}$ definiert wird[4]; jede abgeschlossene Teilmenge eines m-Raumes[5]; usw.

Der zweite Begriff, den wir einführen, ist derjenige eines dem Lebesgueschen äußeren Maß analogen Maßes, definiert in einem m-Raume. Wir nennen es ein l-Maß[6]:

---

[1] Received March 16, 1932.

[2] Vgl. z. B. Hausdorff „Mengenlehre" (Berlin und Leipzig, 1927), S. 94, 125 und 103.

[3] Im ∞-dimensionalen bestehen, wie man sieht, mehrere, voneinander wesentlich verschiedene, Analoga des Euklidischen Raumes!

[4] Diese Entfernungsdefinition hat die Wirkung, daß die Folge $\{\xi_1^{(p)}, \xi_2^{(p)}, \cdots\}$ $(p=1,2,\cdots)$ dann und nur dann für $p\to\infty$ gegen $\{\xi_1, \xi_2, \cdots\}$ konvergiert, wenn für jedes n $\xi_n^{(p)}\to\xi_n$ $(p\to\infty)$ gilt.

[5] Mit derselben Metrik!

[6] Die nachfolgenden Begriffsbildungen lehnen an jene von Carathéodorys Theorie der regulären Maßfunktionen an, jedoch mit einigen Zusatzbedingungen (d. s. Verschärfungen in I. und V.). Vgl. Carathéodory „Reelle Funktionen" (Berlin und Leipzig 1918), S. 237—274.

DEFINITION 2. *Im m-Raume Ω ist ein l-Maß definiert, wenn jeder Teilmenge N von Ω eine Zahl $\mu^*(N)$ zugeordnet ist, mit den folgenden Eigenschaften:*

I. *$\mu^*(N)$ ist 0, oder positiv-endlich, oder $+\infty$. Für jede Kugel ist es endlich. (Eine Kugel mit dem Mittelpunkt x_0 und dem Radius $\varrho>0$ ist die Menge der x mit $\overline{x_0, x} \leqq \varrho$.)*

II. *Aus $M \subset N$[7] folgt $\mu^*(M) \leqq \mu^*(N)$.*

III. *Seien $N_1, N_2, \cdots$ endlich oder abzählbar viele Mengen, M ihre Vereinigungsmenge. Dann ist: $\mu^*(M) \leqq \mu^*(N_1)+\mu^*(N_2)+\cdots$.*

IV. *Ist die Entfernung der Mengen M, N[8] >0, so ist $\mu^*(M+N)=\mu^*(M)+\mu^*(N)$.*

V. *$\mu^*(M)$ ist die untere Grenze aller $\mu^*(O)$, wenn O sämtliche offenen Mengen $\supset M$ durchläuft*[9].

Wir können nun, in beinahe wörtlicher Wiederholung von Carathéodorys Begriffsbildungen und Beweisen (a. a. O. Anmerkung [6]) zunächst definieren:

DEFINITION 3. *Die Teilmenge M von Ω heißt meßbar, wenn für jede Teilmenge N von Ω*

$$\mu^*(N) = \mu^*(MN)+\mu^*(N-MN)$$[10]

gilt. Statt $\mu^*(M)$ schreiben wir dann $\mu(M)$ und beweisen dann:

a) *jedes offene M ist meßbar.*

b) *Sind $M_1, M_2, \cdots$ endlich oder abzählbar viele meßbare Mengen, so ist $M_1+M_2+\cdots$ und $M_1 \cdot M_2 \cdot \cdots$ meßbar. Ferner ist mit M, N auch $M-N$ meßbar.*

c) *Alle M mit $\mu^*(M)=0$ sind meßbar.*

---

[7] D. h.: M ist Teilmenge von N.

[8] D. h. die untere Grenze aller $\overline{x, y}$, x in M, y in N.

[9] II. folgt aus V., wir wollen aber die Carathéodorysche Anordnung nicht ändern. — Übrigens könnten wir in V. an Stelle der offenen Mengen $\supset M$ ebensogut alle Borelschen Mengen $\supset M$ treten lassen. Da jede offene Menge Borelsch ist, ist bloß zu zeigen: wenn P Borelsch ist, so existiert ein offenes $O \supset P$ mit $\mu^*(O)<\mu^*(P)+\varepsilon$ ($\varepsilon>0$). Bilden wir hierzu die Borelschen Klassen $F^{(\xi)}$ nach Hausdorff, S. 178, aber so, daß $F^{(0)}$ aus den offenen Mengen besteht. Ein Borelsches O obiger Art existiert: etwa $O=P$, wählen wir es aus $F^{(\xi)}$ mit kleinstmöglicher Ordnungszahl ξ. Wäre ξ gerade, $\neq 0$, so wäre $O=O_1 \cdot O_2 \cdot \cdots$, O_n aus $F^{(\xi_n)}$, $\xi_n<\xi$, und nach e) im Text $\lim\limits_{n\to\infty} \mu^*(O_n)=\mu^*(O)$, also ein $\mu^*(O_n)<\mu(P)+\varepsilon$, entgegen der Annahme. Wäre ξ ungerade, also $=\eta+1$, so ist $O=O_1+O_2+\cdots$, O_n in $F^{(\eta)}$. Da bei unserer Definition $F^{(1)}=F^{(0)}$ ist, ist $\eta \neq 0$, also $O_n=O_{n1} \cdot O_{n2} \cdot \cdots$, O_{nm} aus $F^{(\eta_{nm})}$, $\eta_{nm}<\eta$. Also ist $\lim\limits_{n\to\infty} \mu^*(O_{nm})=\mu^*(O_n)$, und wenn wir $m=m_n$ mit $\mu^*(O_{nm})<\mu^*(O_n)+\frac{\delta}{2^n}$ ($\delta>0$) wählen, für $O'=O_{1m_1}+O_{2m_2}+\cdots$ $\mu^*(O')<\mu^*(O)+\delta=\mu^*(P)+\varepsilon$, wenn wir $\delta=\mu^*(P)+\varepsilon-\mu^*(O)$ wählen. Dabei gehört O' zu $F^{(\eta)}$, entgegen der Annahme. Also muß $\xi=0$ sein, d. h. O offen.

[10] $M+N$, MN oder $M \cdot N$, $M-N$ bezeichnen die Vereinigungs-, Durchschnitts-, Differenzmenge. Ebenso für mehrere Addenden bzw. Faktoren.

Ferner die bekannten Additivitäts- und Limeseigenschaften des Lebesgueschen Maßes:

d) *Sind* $M_1, M_2, \cdots$ *wie in* b) *und paarweise elementfremd, so ist* $\mu(M_1+M_2+\cdots)=\mu(M_1)+\mu(M_2)+\cdots$.

e) *Sind* $M_1, M_2, \cdots$ *wie in* b) *und* $M_1 \supset M_2 \supset \cdots$ *oder* $M_1 \subset M_2 \subset \cdots$ *und* N *ihr Limes (d. i.* $M_1 \cdot M_2 \cdot \cdots$ *bzw.* $M_1+M_2+\cdots$*), so ist* $\mu(N)=\lim_{n\to\infty}\mu(M_n)$.

Aus a), b) folgt insbesondere, daß alle Borelschen Mengen meßbar sind.

Beispiele von l-Maßen sind leicht anzugeben: das gewöhnliche Lebesguesche äußere Maß im k-dimensionalen Euklidischen Raume; wenn $f(x_1, \cdots, x_k)$ eine in diesem Raume definierte, im gewöhnlichen Lebesgueschen Sinne meßbare Funktion ist, mit Werten >0, die Größe $\mu^*(N)=\int\limits_{\{x_1,\cdots,x_k\}\text{ in }N}\cdots\int f(x_1,\cdots,x_k)\,dx_1\cdots dx_k$[11]; usw. Der allereinfachste Fall besteht, wenn Ω ein Intervall $0\leqq x<a$ ist (a positiv-endlich oder ∞), und μ^* das Lebesguesche äußere Maß ist.

Ist ein m-Raum Ω mit einem l-Maß gegeben, so betrachten wir seine meßbaren Teilmengen. Zwei solche, M, N, nennen wir äquivalent, in Zeichen $M\sim N$, wenn sie sich nur um eine Menge vom Maße 0 unterscheiden, d. h. wenn $\mu^*((M+N)-M\cdot N)=0$ ist. Da offenbar $M\sim M$ gilt, aus $M\sim N$ $N\sim M$ folgt, und aus $M\sim N$, $N\sim \mathfrak{A}$ $M\sim \mathfrak{A}$ folgt, hat dieser Begriff die wesentlichen Eigenschaften eines Äquivalenzbegriffes. Wir definieren nun:

DEFINITION 4. *Seien* Ω, Ω' *zwei m-Räume mit den bzw. l-Maßen* μ^*, μ'^*. *Eine maßtreue Punktabbildung von* Ω *auf* Ω' *ist eine eineindeutige Abbildung* $x'=\varphi(x)$ *von* Ω *auf* Ω', *die jeder meßbaren Teilmenge* M *von* Ω *ein meßbares Bild* M' *in* Ω' *vom gleichen Maße zuordnet, und jeder meßbaren Teilmenge* M' *von* Ω' *ein meßbares Urbild* M *in* Ω *vom gleichen Maße.*

DEFINITION 5. *Seien* Ω, Ω', μ^*, μ'^* *wie vorhin. Eine maßtreue Mengenabbildung von* Ω *auf* Ω' *ist eine solche, die jeder meßbaren Teilmenge von* Ω *eine meßbare Teilmenge von* Ω' *vom gleichen Maße zuordnet, und zwar mit den folgenden Eigenschaften:*

1. *Zu jedem meßbaren* M' *in* Ω' *existiert ein meßbares* M *in* Ω, *dem eine* M' *äquivalente Menge zugeordnet ist.*
2. *Sind* $M_1, M_2, \cdots$ *(in* Ω*) die bzw. Mengen* $M_1', M_2', \cdots$ *(in* Ω'*) zugeordnet, so sind* $M_1+M_2+\cdots$, $M_1\cdot M_2\cdot\cdots$ *bzw.* $M_1'+M_2'+\cdots$, $M_1'\cdot M_2'\cdot\cdots$ *zugeordnet. Sind* M, N *(in* Ω*) bzw.* M', N' *(in* Ω'*) zugeordnet, so ist* $M-N$ $M'-N'$ *zugeordnet*[12].

[11] Falls N nach Lebesgue nicht meßbar ist, bedeute $\int\limits_{\{x_1,\cdots,x_k\}\text{ in }N}\cdots\int$ das „obere" Lebesguesche Integral. $f(x_1,\cdots,x_k)\equiv 1$ führt zum gewöhnlichen Lebesgueschen Maß zurück.

[12] Von diesen drei Forderungen genügen zwei, da aus ihnen die dritte folgt.

Zwei maßtreue Mengenabbildungen (von Ω auf Ω') heißen äquivalent, wenn für jedes meßbare M in Ω die durch die beiden Mengenabbildungen ihm zugeordneten Mengen in Ω' einander äquivalent sind.

Eine maßtreue Punktabbildung $x' = \varphi(x)$ erzeugt offenbar eine maßtreue Mengenabbildung: indem wir jedem meßbaren M in Ω sein durch $x' = \varphi(x)$ vermitteltes Bild M' in Ω' zuordnen[13]. Wir werden aber auch die Umkehrung hiervon beweisen: (Satz 1) zu jeder maßtreuen Mengenabbildung kann eine maßtreue Punktabbildung gefunden werden, die eine ihr äquivalente maßtreue Mengenabbildung erzeugt.

Diese, und noch einige weitere Sätze bilden den Inhalt dieser Arbeit.

2. Seien Ω, Ω' m-Räume mit den bzw. l-Maßen μ^*, μ'^*, und sei eine maßtreue Mengenabbildung von Ω auf Ω' (im Sinne von Def. 5) gegeben. Ehe wir den angekündigten Satz 1 beweisen, leiten wir den folgenden Hilfssatz ab:

HILFSSATZ. Sei M eine meßbare Menge in Ω, und es sei ein $\varepsilon > 0$ gegeben. Dann kann M in paarweise elementfremde Mengen $\Gamma_1, \Gamma_2, \cdots, N$ zerlegt werden, $M = \Gamma_1 + \Gamma_2 + \cdots + N$ (die Anzahl der Addenden Γ_n darf auch endlich, oder sogar 0, sein), derart, daß $\mu(N) = 0$ ist, und jedes Γ_n abgeschlossen ist, $\mu(\Gamma_n) > 0$, und einen Diameter $\leqq \varepsilon$ hat[14].

Die Bedingung über den Diameter ist offenbar gegenstandslos, wenn M selbst einen Diameter $\leqq \varepsilon$ hat. Ist ferner $M = M_1 + M_2 + \cdots$, wo $M_1, M_2, \cdots$ paarweise elementfremde meßbare Mengen sind, für die der Hilfssatz gesichert ist, so steht er auch für M fest: denn

$$M_1 = \Gamma_{11} + \Gamma_{12} + \cdots + N_1, \qquad M_2 = \Gamma_{21} + \Gamma_{22} + \cdots + N_2, \cdots$$

ergibt für M die Zerlegung

$$M = \Gamma_1 + \Gamma_2 + \cdots + N,$$

wo $\Gamma_1, \Gamma_2, \cdots$ die $\Gamma_{11}, \Gamma_{12}, \cdots, \Gamma_{21}, \Gamma_{22}, \cdots, \cdots$ in irgendeiner Reihenfolge sind, und $N = N_1 + N_2 + \cdots$. Es genügt also, den Hilfssatz ohne die Diameterbedingung zu beweisen, wenn jedes meßbare $M = M_1 + M_2 + \cdots$, mit paarweise elementfremden, meßbaren $M_1, M_2, \cdots$, mit Diametern $\leqq \varepsilon$ ist. Nun ist Ω separabel, sei $x_1, x_2, \cdots$ eine überall dichte Folge, $K_1, K_2, \cdots$ die Kugeln vom Radius ε um diese bez. Mittelpunkte. Dann ist

$$\Omega = K_1 + K_2 + \cdots,$$

so daß

$$M_1 = K_1 \cdot M, \qquad M_2 = (K_2 - K_1 \cdot K_2) \cdot M,$$
$$M_3 = (K_3 - (K_1 + K_2) \cdot K_3) \cdot M, \cdots$$

das Gewünschte leisten.

[13] 1. gilt sogar in verschärfter Form: jedes meßbare M' in Ω' ist selbst einem meßbaren M in Ω (seinem durch $x' = \varphi(x)$ vermittelten Urbild) zugeordnet.

[14] Der Diameter einer Menge Γ ist die obere Grenze aller Zahlen $\overline{x, y}$, x in Γ, y in Γ.

Wir können also die Diameterbedingung fortlassen, und M in einer Kugel K gelegen annehmen. K ist abgeschlossen, also meßbar, also $K-M$ auch. Sei $\delta>0$, nach Def. 2, V existiert ein offenes $O\supset K-M$, mit $\mu(O)\leqq\mu(K-M)+\delta$. Also ist erst recht

$$\mu(O\cdot K)\leqq\mu(K-M)+\delta,$$

und

$$\mu(K-O\cdot K)=\mu(K)-\mu(O\cdot K)\geqq\mu(K)-\mu(K-M)-\delta=\mu(M)-\delta.$$

Dabei ist $\Gamma=K-O\cdot K$ abgeschlossen (K ist abgeschlossen, O offen) und $\subset M$.

Bilden wir nun das Γ für M und $\delta=\frac{1}{1}$, es heiße Γ_1, dann das Γ für $M-\Gamma_1$ und $\delta=\frac{1}{2}$, es heiße Γ_2, dann das Γ für $M-\Gamma_1-\Gamma_2$ und $\delta=\frac{1}{3}$, es heiße $\Gamma_3,\cdots$. Nach Konstruktion ist

$$\mu(M-\Gamma_1-\cdots-\Gamma_n)\leqq\frac{1}{n+1},$$

also auch

$$\mu(M-\Gamma_1-\Gamma_2-\cdots)\leqq\frac{1}{n+1},\quad\text{d. h.}=0.$$

Diese $\Gamma_1,\Gamma_2,\cdots$ und $N=M-\Gamma_1-\Gamma_2-\cdots$ leisten daher alles Gewünschte, wenn wir noch diejenigen unter ihnen, die das Maß 0 haben, falls solche überhaupt vorliegen, fortlassen und zu N hinzufügen. Damit ist der Hilfssatz bewiesen.

Nunmehr gehen wir so vor: Wir zerlegen Ω nach dem Hilfssatz in $\Gamma_1+\Gamma_2+\cdots+N$ mit $\varepsilon=\frac{1}{1}$. Die Mengenabbildung ordne $\Gamma_1,\Gamma_2,\cdots,N$ die Mengen $\Gamma_1',\Gamma_2',\cdots,N'$ in Ω' zu. Je zwei der ersteren Mengen haben leere Durchschnitte, also haben je zwei der letzteren Durchschnitte vom Maße 0. Daher sind $\Gamma_1',\Gamma_2',\Gamma_3',\cdots$ bzw. den Mengen

$$\overline{\Gamma}_1'=\Gamma_1',\qquad\overline{\Gamma}_2'=\Gamma_2'-\Gamma_1'\cdot\Gamma_2',\qquad\overline{\Gamma}_3'=\Gamma_3'-(\Gamma_1'+\Gamma_2')\cdot\Gamma_3',\cdots$$

äquivalent. Also: $\Gamma_1,\Gamma_2,\cdots$ sind paarweise elementfremd und $\subset\Omega$, ebenso $\overline{\Gamma}_1',\overline{\Gamma}_2',\cdots$ und $\subset\Omega'$; $\Gamma_1+\Gamma_2+\cdots$ ist $\subset\Omega$ und ihm äquivalent, $\overline{\Gamma}_1'+\overline{\Gamma}_2'+\cdots$ ist $\subset\Omega'$ und ihm äquivalent; die Γ_n zugeordnete Menge ist $\overline{\Gamma}_n'$ äquivalent; Γ_n ist abgeschlossen, vom Maße >0, und sein Diameter $\leqq\frac{1}{1}$.

Jetzt zerlegen wir jedes $\overline{\Gamma}_n'$ nach dem Hilfssatz in $\Gamma_{n1}'+\Gamma_{n2}'+\cdots+N_n'$ mit $\varepsilon=\frac{1}{1}$. Die Mengenabbildung ordne $\Gamma_{n1}',\Gamma_{n2}',\cdots,N_n'$ die Mengen $\Gamma_{n1},\Gamma_{n2},\cdots,N_n$ in Ω zu. Da je zwei der ersteren Mengen leere Durchschnitte haben und $\subset\overline{\Gamma}_n'$ sind, also $\subset$ in einer Menge $\sim\Gamma_n'$, können wir die letzteren wiederum durch äquivalente Mengen ersetzen, die auch paarweise elementfremd und $\subset\Gamma_n$ sind. So werde aus $\Gamma_{n1},\Gamma_{n2},\cdots$ bzw. $\overline{\Gamma}_{n1},\overline{\Gamma}_{n2},\cdots$. Also: $\overline{\Gamma}_{n1},\overline{\Gamma}_{n2},\cdots$ sind paarweise elementfremd und $\subset\Gamma_n$, ebenso Γ_{n1}',

$\Gamma'_{n2}, \cdots$ und $\subset \bar{\Gamma}'_n$; $\bar{\Gamma}_{n1} + \bar{\Gamma}_{n2} + \cdots$ ist $\subset \Gamma_n$ und ihm äquivalent, $\Gamma'_{n1} + \Gamma'_{n2} + \cdots$ ist $\subset \bar{\Gamma}'_n$ und ihm äquivalent; die $\bar{\Gamma}_{np}$ zugeordnete Menge ist Γ'_{np} äquivalent; Γ'_{np} ist abgeschlossen, vom Maße >0, und sein Diameter $\leqq \frac{1}{1}$.

Die mit Ω, Ω' und $\frac{1}{1}$ vorgenommene Konstruktion wiederholen wir an $\bar{\Gamma}_{np}$, Γ'_{np} und $\frac{1}{2}$. So entstehen Mengen Γ_{npq}, $\bar{\Gamma}'_{npq}$ und $\bar{\Gamma}_{npqr}$, Γ'_{npqr} mit den folgenden Eigenschaften: $\Gamma_{np1}, \Gamma_{np2}, \cdots$ sind paarweise elementfremd und $\subset \bar{\Gamma}_{np}$, ebenso $\bar{\Gamma}'_{np1}, \bar{\Gamma}'_{np2}, \cdots$ und $\subset \Gamma'_{np}$; $\Gamma_{np1} + \Gamma_{np2} + \cdots$ ist $\subset \bar{\Gamma}_{np}$ und ihm äquivalent, $\bar{\Gamma}'_{np1}, \bar{\Gamma}'_{np2}, \cdots$ ist $\subset \Gamma'_{np}$ und ihm äquivalent; die Γ_{npq} zugeordnete Menge ist $\bar{\Gamma}'_{npq}$ äquivalent; Γ_{npq} ist abgeschlossen, vom Maße >0, und sein Diameter $\leqq \frac{1}{2}$. $\bar{\Gamma}_{npq1}, \bar{\Gamma}_{npq2}, \cdots$ sind paarweise elementfremd und $\subset \Gamma_{npq}$, ebenso $\Gamma'_{npq1}, \Gamma'_{npq2}, \cdots$ und $\subset \bar{\Gamma}'_{npq}$; $\bar{\Gamma}_{npq1} + \bar{\Gamma}_{npq2} + \cdots$ ist $\subset \Gamma_{npq}$ und ihm äquivalent, $\Gamma'_{npq1} + \Gamma'_{npq2} + \cdots$ ist $\subset \bar{\Gamma}'_{npq}$ und ihm äquivalent; die $\bar{\Gamma}_{npqr}$ zugeordnete Menge ist Γ'_{npqr} äquivalent; Γ'_{npqr} ist abgeschlossen, vom Maße >0, und sein Diameter $\leqq \frac{1}{2}$.

Dieselbe Konstruktion wiederholen wir an $\bar{\Gamma}_{npqr}$, Γ'_{npqr} und $\frac{1}{3}$, und gewinnen dadurch Mengen Γ_{npqrs}, $\bar{\Gamma}'_{npqrs}$ und $\bar{\Gamma}_{npqrst}$, Γ'_{npqrst} usw. Zusammenfassend entsteht ein System von Mengen $\Gamma_{np\ldots fg}$, $\bar{\Gamma}'_{np\ldots fg}$ und $\bar{\Gamma}_{np\ldots fgh}$, $\Gamma'_{np\ldots fgh}$ ($n, p, \cdots, f, g$ sei eine beliebige ungerade Anzahl von Indizes, etwa $2\nu-1$, deren jeder über $1, 2, \cdots$ oder ein Anfangsstück davon läuft, $n, p, \cdots, f, g, h$ entsprechend eine gerade Anzahl 2ν). Dabei gilt: $\Gamma_{np\ldots f1}, \Gamma_{np\ldots f2}, \cdots$ paarweise elementfremd und $\subset \bar{\Gamma}_{np\ldots f}$, ebenso $\bar{\Gamma}'_{np\ldots f1}$, $\bar{\Gamma}'_{np\ldots f2}, \cdots$ und $\subset \Gamma'_{np\ldots f}$; $\Gamma_{np\ldots f1} + \Gamma_{np\ldots f2} + \cdots$ ist $\subset \bar{\Gamma}_{np\ldots f}$ und ist ihm äquivalent, $\bar{\Gamma}'_{np\ldots f1} + \bar{\Gamma}'_{np\ldots f2} + \cdots$ ist $\subset \Gamma'_{np\ldots f}$ und ist ihm äquivalent; die $\Gamma_{np\ldots fg}$ zugeordnete Menge ist $\bar{\Gamma}'_{np\ldots fg}$ äquivalent; $\Gamma_{np\ldots fg}$ ist abgeschlossen, vom Maße >0, und sein Diameter ist $\leqq \frac{1}{\nu}$. Analoges gilt für $\bar{\Gamma}_{np\ldots fgh}$, $\Gamma'_{np\ldots fgh}$, nur ist dort $\Gamma'_{np\ldots fgh}$ abgeschlossen, vom Maße >0, und vom Diameter $\leqq \frac{1}{\nu}$. Für $\nu = 1$ tritt an Stelle von $\Gamma_{np\ldots f}$, $\bar{\Gamma}'_{np\ldots f}$ (das $2\nu - 2 = 0$ Indizes hätte) Ω, Ω'.

Alle Mengen $\bar{\Gamma}_{np\ldots f} - \Gamma_{np\ldots f1} - \Gamma_{np\ldots f2} - \cdots$, $\Gamma'_{np\ldots f} - \bar{\Gamma}'_{np\ldots f1} - \bar{\Gamma}'_{np\ldots f2} - \cdots$, $\Gamma_{np\ldots fg} - \bar{\Gamma}_{np\ldots fg1} - \bar{\Gamma}_{np\ldots fg2} - \cdots$, $\bar{\Gamma}'_{np\ldots fg} - \Gamma'_{np\ldots fg1} - \Gamma'_{np\ldots fg2} - \cdots$ haben das Maß 0, also auch ihre Vereinigungsmengen über alle ν und $n, p, \cdots, f, g, h$ (der ersten und dritten in Ω, der zweiten und vierten in Ω'): M_0, M'_0. Aus dem bisher Gesagten folgt, daß jeder Punkt von $\Omega - M_0$ genau einem Durchschnitt $\Gamma_n \cdot \bar{\Gamma}_{np} \cdot \Gamma_{npq} \cdot \bar{\Gamma}_{npqr} \cdot \cdots$ angehört, und jeder Punkt von $\Omega' - M'_0$ genau einem Durchschnitt $\bar{\Gamma}'_n \cdot \Gamma'_{np} \cdot \bar{\Gamma}'_{npq} \cdot \Gamma'_{npqr} \cdot \cdots$. Andererseits ist jeder Durchschnitt $\Gamma_n \cdot \bar{\Gamma}_{np} \cdot \Gamma_{npq} \cdot \bar{\Gamma}_{npqr} \cdot \cdots$ wegen $\Gamma_n \supset \bar{\Gamma}_{np} \supset \Gamma_{npq} \supset \bar{\Gamma}_{npqr} \supset \cdots$ gleich $\Gamma_n \cdot \Gamma_{npq} \cdot \cdots$. Da Γ_n, Γ_{npq}, ... eine absteigende Folge nicht leerer abgeschlossener Mengen ist, deren Diameter gegen 0 streben, besteht ihr Durchschnitt aus genau einem Punkt (vgl. a. a. O. Anm. [2]). Ebenso zeigt

man, daß der Durchschnitt $\overline{\Gamma}'_n \cdot \Gamma'_{np} \cdot \overline{\Gamma}'_{npq} \cdot \Gamma'_{npqr} \cdots$ aus genau einem Punkte besteht (hier ist $\Gamma'_{np} \cdot \Gamma'_{npqr} \cdots$ heranzuziehen). Sei die Menge aller $\Gamma_n \cdot \overline{\Gamma}_{np} \cdot \Gamma_{npq} \cdot \overline{\Gamma}_{npqr} \cdots$-Durchschnittspunkte Ω_1, die aller $\overline{\Gamma}'_n \cdot \Gamma'_{np} \cdot \overline{\Gamma}'_{npq} \cdot \Gamma'_{npqr} \cdots$-Durchschnittspunkte Ω'_1. Da $\Omega_1 \supset \Omega - M_0$, $\Omega - \Omega_1 \subset M_0$ ist, ist $\Omega - \Omega_1$ vom Maße 0, d. h. Ω_1 dem Ω äquivalent, ebenso ist Ω'_1 dem Ω' äquivalent.

Auf diese Weise sind sowohl die Punkte von Ω_1, als auch diejenigen von Ω'_1, eineindeutig auf die Folgen $n, p, q, r, \cdots$ abgebildet; dies bewirkt auch eine eineindeutige Abbildung von Ω_1 auf Ω'_1, welche $x' = \varphi(x)$ heißen möge.

Aus der Definition dieser Abbildung folgt, daß sie $\Gamma_{np\ldots fg} \cdot \Omega_1$ in $\overline{\Gamma}'_{np\ldots fg} \cdot \Omega'_1$ überführt. Diese Mengen sind $\Gamma_{np\ldots fg}$ bzw. $\overline{\Gamma}'_{np\ldots fg}$ äquivalent, haben also bzw. dieselben Maße wie diese. $\Gamma_{np\ldots fg}$ hat dasselbe Maß, wie die ihr zugeordnete Menge, diese wieder dasselbe Maß, wie das ihr äquivalente $\overline{\Gamma}'_{np\ldots fg}$. Also haben $\Gamma_{np\ldots fg} \cdot \Omega_1$ und $\overline{\Gamma}'_{np\ldots fg} \cdot \Omega'_1$, ihr Bild, dasselbe Maß. Wenn Δ die Summe einiger Mengen $\Gamma_{np\ldots fg}$ ist (alle von gleicher Indizesanzahl $2\nu - 1$, also paarweise elementfremd), so ist demnach das $x' = \varphi(x)$-Bild von $\Delta \cdot \Omega_1$ meßbar und vom selben Maß wie diese Menge.

Sei M abgeschlossen (in Ω). $\Gamma^{(\nu)}$ sei die Summe aller $\Gamma_{np\ldots fg}$ mit $2\nu - 1$ Indizes, die Punkte aus M enthalten. Es ist $\Gamma^{(\nu)} \supset M$, und da jedes $\Gamma_{np\ldots fg}$ einen Diameter $\leqq \frac{1}{\nu}$ hat, hat jeder Punkt von $\Gamma^{(\nu)}$ eine Distanz $\leqq \frac{1}{\nu}$ von M. Daher ist $M = \Gamma^{(1)} \cdot \Gamma^{(2)} \cdots$, ferner $\Gamma^{(1)} \supset \Gamma^{(2)} \supset \cdots$; also auch

$$M \cdot \Omega_1 = (\Gamma^{(1)} \cdot \Omega_1) \cdot (\Gamma^{(2)} \cdot \Omega_1) \cdots, \qquad \Gamma^{(1)}\Omega_1 \supset \Gamma^{(2)}\Omega_1 \supset \cdots.$$

Da das Bild von $\Gamma^{(\nu)} \cdot \Omega_1$ meßbar und vom selben Maße ist, gilt dasselbe vom Bilde von $M \cdot \Omega_1$.

Wenn O offen ist, so ist $O = M_1 + M_2 + \cdots$, wobei $M_1, M_2, \cdots$ abgeschlossen und $M_1 \subset M_2 \subset \cdots$ sind; also $O \cdot \Omega_1 = M_1 \cdot \Omega_1 + M_2 \cdot \Omega_1 + \cdots$, $M_1 \cdot \Omega_1 \subset M_2 \cdot \Omega_1 \subset \cdots$, und da das Bild von $M_n \cdot \Omega_1$ meßbar und vom selben Maße ist, gilt dasselbe von $O \cdot \Omega_1$. Ist $N = O_1 \cdot O_2 \cdots$, wobei $O_1, O_2, \cdots$ offen und $O_1 \supset O_2 \supset \cdots$ sind, so gilt dasselbe wegen $N \cdot \Omega_1 = (O_1 \cdot \Omega_1) \cdot (O_2 \cdot \Omega_1) \cdots$, $O_1 \cdot \Omega_1 \supset O_2 \cdot \Omega_1 \supset \cdots$, auch für $N \cdot \Omega_1$. Schließlich ist $O_1 \supset O_2 \supset \cdots$ überflüssig, da wir $O_1, O_2, \cdots$ durch $O_1, O_1 \cdot O_2, \cdots$ ersetzen können.

Wenn A das Maß 0 hat, so gibt es ein $M = O_1 \cdot O_2 \cdots \supset A$ vom Maß 0 (wegen Def. 2, V). Das Bild von $A \cdot \Omega_1$ ist Teil des Bildes von $M \cdot \Omega_1$, da dieses das Maß 0 hat, hat jenes auch das Maß 0 und ist meßbar.

Die Aussage: das Bild von $M \cdot \Omega_1$ ist meßbar und hat dasselbe Maß wie M, gilt also 1. für jedes $M = O_1 \cdot O_2 \cdots$, $O_1, O_2, \cdots$ offen; 2. für jedes M vom Maße 0. Da jedes meßbare M gleich einem von der Art 1. minus einem von der Art 2. ist (wegen Def. 2, V, oder durch Anwendung des Hilfssatzes auf $\Omega - M$), gilt die obige Aussage für jedes meßbare M.

Also wird jedes meßbare $M \subset \Omega_1$ durch $x' = \varphi(x)$ auf ein meßbares $M' \subset \Omega_1'$ vom selben Maße abgebildet. Auf dieselbe Weise beweist man die Umkehrung. Nach Def. 1 bildet also $x' = \varphi(x)$ Ω_1 auf Ω_1' maßtreu ab.

Für welche M ist das $x' = \varphi(x)$-Bild von $M \cdot \Omega_1$ der zugeordneten Menge (in Ω') äquivalent? Für $M = \Gamma_{np\ldots fg}$ ist das jedenfalls der Fall (das Bild ist $\overline{\Gamma}'_{np\ldots fg} \cdot \Omega_1'$, äquivalent $\overline{\Gamma}'_{np\ldots fg}$, also der Menge, die $\Gamma_{np\ldots fg}$ zugeordnet ist); also auch für die weiter oben erwähnten Δ; also bei abgeschlossenem M für die weiter oben erwähnten $\Gamma^{(\nu)}$, und für $M = \Gamma^{(1)} \cdot \Gamma^{(2)} \cdot \ldots$ selbst; also, auf Grund derselben Betrachtungen wie weiter oben nacheinander 1. für jedes offene O, 2. für jedes $M = O_1 \cdot O_2 \cdot \ldots$, O_1, O_2, $\ldots$ offen, 3. für jedes M vom Maße 0, 4. für jedes meßbare M. Somit ist die durch $x' = \varphi(x)$ erzeugte Mengenzuordnung der ursprünglich gegebenen äquivalent.

Ein Mangel besteht noch: nämlich daß $x' = \varphi(x)$ Ω_1 auf Ω_1' abbildet, und nicht Ω auf Ω'. Dieser wird, falls Ω, Ω' unabzählbar sind, wie folgt behoben. Sei $M_1 \subset \Omega$ vom Maße 0 und von der Mächtigkeit des Kontinuums,[15] $M_1' \subset \Omega'$ ebenso. $M_1' \cdot \Omega_1'$ hat auch das Maß 0, also auch sein Urbild M_2, ebenso $M_1 \cdot \Omega_1$ und sein Bild M_2'. Somit sind $M_1 + M_2 \subset \Omega$ und $M_1' + M_2' \subset \Omega'$ vom Maße 0 und der Mächtigkeit des Kontinuums,[16] und $(M_1 + M_2) \cdot \Omega_1$ hat das Bild $(M_1' + M_2') \cdot \Omega_1'$. Wir setzen

$$\Omega_2 = \Omega_1 - (M_1 + M_2) \cdot \Omega_1, \qquad \Omega_2' = \Omega_2 - (M_1' + M_2') \cdot \Omega_2,$$

dann ist Ω_2' das Bild von Ω_2, und da $\Omega_1 - \Omega_2 \subset M_1 + M_2$ das Maß 0 hat, hat auch $\Omega - \Omega_2$ das Maß 0, und weil es $\supset M_1 + M_2$ ist, die Mächtigkeit des Kontinuums. Dasselbe gilt für $\Omega' - \Omega_2'$.

Schränken wir nun die Abbildung $x' = \varphi(x)$ auf Ω_2, Ω_2' ein, verwenden wir dagegen in $\Omega - \Omega_2$, $\Omega' - \Omega_2'$ irgendeine eineindeutige Abbildung $x' = \psi(x)$ der ersteren Menge auf die letztere (beide haben ja dieselbe Mächtigkeit). Beide Abbildungen machen zusammen eine eineindeutige Abbildung $x' = \chi(x)$ von Ω auf Ω' aus. Da sich aber $x' = \chi(x)$ von $x' = \varphi(x)$ (in Ω wie in Ω') nur auf einer Menge vom Maße 0 unterscheidet, hat die erstere

[15] Eine solche Menge konstruiert man z. B. so: Nach Hausdorff, S. 138, Satz XI und S. 137, Satz X, hat das unabzählbare Ω ein „dyadisches Diskontinuum" D zur Teilmenge. D. h. es ist jeder Folge $u_1, u_2, \cdots$ $(= 0, 1)$ ein Element $x_{u_1 u_2 \ldots}$ von Ω zugeordnet, derart daß für jede gegebene Umgebung O von $x_{u_1 u_2 \ldots}$ ein n existiert, daß alle $x_{v_1 v_2 \ldots}$ mit $u_1 = v_1, \cdots, u_n = v_n$ zu O gehören; und D die Menge aller $x_{u_1 u_2 \ldots}$ ist. Seien nun $u_1, u_2, \cdots$ fest gegeben, und $D_{u_1 u_2 \ldots}$ die Menge aller $x_{u_1 u_2 \ldots}$ ($u_2, u_4, \cdots$ beliebig). Dabei ist $D \subset K$ für eine geeignete Kugel K. Die $D_{u_1 u_2 \ldots}$ sind alle abgeschlossen, von der Mächtigkeit des Kontinuums, und $\subset K$. Je zwei solche Mengen sind elementfremd, also ist die Maßsumme irgendwelcher endlich vielen unter ihnen $\leqq \mu(K)$ (endlich!). Daher können höchstens abzählbar viele unter ihnen Maße > 0 haben. Aber ihre Anzahl ist Kontinuum, also gibt es unter ihnen auch solche vom Maße 0.

[16] Ω, Ω' haben selbst, als separable Mengen, höchstens die Mächtigkeit des Kontinuums.

Abbildung mit der letzteren dies gemein: sie ist maßtreu, und erzeugt eine Mengenzuordnung, welche der ursprünglich gegebenen äquivalent ist.

Damit haben wir bewiesen:

SATZ 1. *Seien Ω, Ω' m-Räume mit den bzw. l-Maßen μ^*, μ'^*. Jede maßtreue Punktabbildung von Ω auf Ω', oder auch nur von Ω_1 auf Ω_1' mit $\Omega_1 \subset \Omega$, $\Omega_1' \subset \Omega'$, $\Omega - \Omega_1$ und $\Omega' - \Omega_1'$ vom Maße 0, erzeugt eine maßtreue Mengenabbildung von Ω auf Ω'. Umgekehrt ist jede maßtreue Mengenabbildung von Ω auf Ω' einer solchen äquivalent, die durch eine maßtreue Punktabbildung eines Ω_1 auf ein Ω_1' (vgl. oben) erzeugt wird. Sind Ω, Ω' unabzählbar, so können wir sogar $\Omega_1 = \Omega$, $\Omega_1' = \Omega$ erreichen. (Vgl. hierzu Def. 4, 5 in § 1, und das daran anschließend Gesagte.)*

3. Seien Ω, Ω', μ^*, μ'^* wie bisher, $x' = \varphi(x)$ eine eineindeutige Abbildung von Ω auf Ω', von der bloß die eine Hälfte der in Def. 4 formulierten Eigenschaften verlangt werde: das Urbild eines jeden meßbaren M' in Ω' soll ein meßbares M in Ω vom selben Maße sein. Wir wollen zeigen, daß dennoch die ganze Def. 4 (d. h. auch die umgekehrte Eigenschaft) gilt, d. h. daß $x' = \varphi(x)$ maßtreu ist.

Sei $x_1', x_2', \cdots$ eine in Ω' überall dichte Folge, $K_{\nu\lambda}'$ $(\nu, \lambda = 1, 2, \cdots)$ die Kugel mit dem Mittelpunkt x_ν' und dem Radius $\frac{1}{\lambda}$. Das Urbild von $K_{\nu\lambda}'$ ist meßbar, also bis auf eine Menge $N_{\nu\lambda}$ vom Maße 0 eine Borelsche Menge (vgl. den Hilfssatz in § 2), diese heiße $A_{\nu\lambda}$. Die Summe aller $N_{\nu\lambda}$ ist auch vom Maße 0, sie ist ~~Teil~~ einer Borelschen Menge M_{00} vom Maße 0 (vgl. die analogen Schlüsse in § 2). In $\Omega - M_{00}$ gilt also: $\varphi(x)$ liegt dann und nur dann in $K_{\nu\lambda}'$, wenn x in $A_{\nu\lambda}$ liegt, d. h. soweit es in $A_{\nu\lambda} - M_{00} \cdot A_{\nu\lambda}$ liegt. Für die auf $\Omega - M_{00}$ eingeschränkte Funktion $\varphi(x)$, sie heiße $\varphi_1(x)$, gilt also: das Urbild von M' (in Ω', M' braucht nicht nur aus $\varphi_1(x)$-Punkten zu bestehen, sein Urbild sind jene x, deren $\varphi_1(x)$ in M' liegen, — übrigens soll im Folgenden der Begriff „Urbild" stets so verstanden werden) ist eine Borelsche Menge, falls M' ein $K_{\nu\lambda}'$ ist. Da jedes offene M' Summe einer Folge von $K_{\nu\lambda}'$'s ist, gilt dies auch für alle offenen M', und dann auch für deren Komplementären, die abgeschlossenen M'. Hieraus wieder schließt man, daß es für alle Borelschen M' gilt.

Sei Ω'' der in § 1 als drittes Beispiel eines m-Raumes erwähnte ∞-dimensionale Raum aller $\{\xi_1, \xi_2, \cdots\}$ (vgl. auch Anm. [4]). In dem wir jedem x' von Ω' das $\psi(x') = \{\overline{x_1', x'}, \overline{x_2', x'}, \cdots\}$ von Ω'' zuordnen, bilden wir Ω' eineindeutig und mitsamt seiner Umkehrung stetig auf eine Teilmenge Ω_0'' von Ω'' ab. In der Tat: Aus $\psi(x'^{(1)}) = \psi(x'^{(2)})$ folgt $\overline{x_n', x'^{(1)}} = \overline{x_n', x'^{(2)}}$, wenn wir also eine Folge $x_{n_1}', x_{n_2}', \cdots$ wählen, die gegen $x'^{(1)}$ strebt, so ist $\overline{x_{n_\nu}', x'^{(1)}} \to 0$, $\overline{x_{n_\nu}', x'^{(2)}} \to 0$, d. h. $x_{n_1}', x_{n_2}', \cdots$ strebt

auch gegen $x'^{(2)}$, $x'^{(1)} = x'^{(2)}$. Aus $x'^{(m)} \to x'$ folgt $\overline{x'_n, x'^{(m)}} \to \overline{x'_n, x}$ (für $m \to \infty$), also $\psi(x'^{(m)}) \to \psi(x')$ (vgl. Anm. [4]). Aus $\psi(x'^{(m)}) \to \psi(x')$, d. h $\overline{x'_n, x'^{(m)}} \to \overline{x'_n, x'}$ für jedes n folgt: sei $\delta > 0$, x'_n mit $\overline{x'_n, x'} \leqq \delta$ gewählt, dann ist für ein genügend großes m $\overline{x'_n, x'^{(m)}} \leqq 2\delta$, also $\overline{x', x'^{(m)}} \leqq \delta + 2\delta = 3\delta$ — somit gilt $x'^{(m)} \to x'$.

Als eineindeutiges stetiges Bild des m-Raumes Ω' ist Ω''_0 (absolut) Borelsch[17]. Sei $M'' \subset \Omega''$ (absolut) Borelsch, dann ist es auch $M'' \cdot \Omega''_0$, das ψ-Urbild von M'' ist das durch die Inverse von ψ vermittelte Bild von $M'' \cdot \Omega''_0$, also wieder (absolut) Borelsch[18]. Das $\psi(\varphi_1)$-Urbild von M'' ist somit das φ_1-Urbild einer Borelschen Menge in Ω', also, wie wir w. o. zeigten, selbst eine Borelsche Menge in Ω. Somit bildet $\psi(\varphi_1(x))$ $\Omega - M_{00}$ eineindeutig auf eine gewisse Teilmenge von Ω'' ab, und jede Borelsche Teilmenge von Ω'' hat als Urbild eine Borelsche Menge in Ω. Daher ist $\psi(\varphi_1(x))$ eine Bairesche Funktion[18]. Infolgedessen hat jedes Borelsche $M \subset \Omega - M_{00}$ ein $\psi(\varphi_1)$-Bild, das ebenfalls Borelsch ist[19], und das ψ-Urbild desselben ist, wie wir weiter oben zeigten, ebenfalls Borelsch. Aber dies ist das φ_1-Bild von M, das mit dessen φ-Bild zusammenfällt, diese sind mithin als Borelsch erwiesen.

Insbesondere ist das Bild des ganzen $\Omega - M_{00}$, es heiße $\Omega' - M'_{00}$, Borelsch, daher ist auch M'_{00} Borelsch und infolgedessen meßbar[20]. Sein Maß ist dasselbe wie dasjenige seines φ-Urbildes M_{00}, also 0. Das φ-Bild M' eines Borelschen M ist die Summe derer von $M - M_{00} \cdot M$ und von $M_{00} \cdot M$. Ersteres ist ein φ_1-Bild, also Borelsch, also meßbar; letzteres $\subset M'_{00}$, also ebenso wie M'_{00} vom Maße 0, also auch meßbar. Daher ist auch M' meßbar, sein Maß ist gleich demjenigen seines Urbildes M.

Ist M vom Maße 0, so ist es $\subset N$, wobei N Borelsch und vom Maße 0 ist (vgl. die entsprechenden Betrachtungen in § 2); das φ-Bild M' von M ist Teil des φ-Bildes N' von N, also wie jenes vom Maße 0, und daher meßbar. Ist schließlich M nur als meßbar vorausgesetzt, so ist es als Summe einer Borelschen Menge und einer vom Maße 0 darstellbar (vgl. den Hilfs-

[17] Vgl. Hausdorff, S. 208—209, insbesondere Satz II.

[18] Vgl. Hausdorff (a. a. O. Anm. [2]), S. 260, Satz V. Allerdings müßte $f(x) = \psi(\varphi_1(x))$ reelle Zahlen als Werte haben, obwohl es in Wahrheit ein $\{\xi_1, \xi_2, \cdots\}$, d. h. $\{\xi_1(x), \xi_2(x), \cdots\}$, ist. Indessen ist der genannte Satz auf jedes $\xi_n(x)$ direkt anwendbar, und überträgt sich von diesen auf $f(x)$ (vgl. Anm. [4]). Mit $\varphi_1(x)$ selbst wäre dieser Schluß nicht möglich gewesen, da es Werte aus Ω' annimmt, das in keiner Beziehung zu den reellen Zahlen steht. Vgl. hierzu a. a. O., S. 268 unten, 269 oben.

[19] Vgl. Hausdorff (a. a. O. Anm. [2]), S. 269, Satz XIII.

[20] Dies ist die Hauptschwierigkeit des Beweises: denn wohl mußte das Urbild N eines meßbaren N' meßbar sein, wir konnten aber aus der Meßbarkeit von N nicht direkt auf diejenige des Bildes N' schließen.

satz in § 2), sein φ-Bild M' ist die Summe der φ-Bilder jener Mengen, also auch meßbar. Das Maß von M' ist jenem seines Urbildes M gleich.

Damit ist die Maßtreue der Abbildung $x' = \varphi(x)$ bewiesen. Wären wir von der anderen Hälfte von Def. 4 ausgegangen, wonach das Bild jedes meßbaren M in Ω ein meßbares M' in Ω' vom selben Maße ist, so wäre dies für die Inverse von φ unsere frühere Prämisse gewesen. Daher hätte sich die Maßtreue dieser Inversen, also auch diejenige von φ, ergeben. Wir haben also bewiesen:

SATZ 2. *Seien Ω, Ω', μ^*, μ'^* wie bisher, $x' = \varphi(x)$ eine eineindeutige Abbildung von Ω auf Ω'. Von den zwei Bedingungen in Def. 4 (Schluß von $M \subset \Omega$ auf $M' \subset \Omega'$, Schluß von $M' \subset \Omega'$ auf $M \subset \Omega$) folgt jede aus der anderen; d. h. jede von ihnen ist auch allein für die Maßtreue von φ notwendig und hinreichend.*

4. Zum Schluß soll ein Satz aus einer früheren Abhandlung des Verfassers[21] verallgemeinert und verschärft werden. Dazu muß eine dortige Begriffsbildung (d. i. die Maßfunktion, vgl. a. a. O., S. 193 unten) verallgemeinert werden.

DEFINITION 6. *Sei Ω ein m-Raum, μ^* ein l-Maß in ihm, u. zw. derart, daß $\mu(\Omega)$ (also jedes $\mu^*(M)$) endlich ist. Eine Funktion $f(x)$, die in Ω definiert ist und reelle Zahlen als Werte hat, heiße meßbar, wenn jedes Intervall $\xi \leqq a$ eine meßbare x-Menge zur $\xi = f(x)$-Urbildmenge hat*[22]. *Die ganze Theorie des Lebesgueschen Integrals kann offenbar, unter Zugrundelegung des Maßes μ^* (bzw. μ), auf diese meßbaren Funktionen f übertragen werden. Das so entstehende Integral bezeichnen wir mit $\int_\Omega f(x)\, dv_x$.*

Ist $f(x)$ meßbar, nie negativ, und beschränkt, so definieren wir für alle a, $0 \leqq a \leqq Max(f(x))$, eine Funktion $\varphi(a) = \mu$ (Menge der x mit $f(x) \leqq a$). $\varphi(a)$ ist monoton nichtfallend und nach rechts halbstetig. Ferner ist

$$0 \leqq \varphi(a) \leqq Max(\varphi(a)) = \mu(\Omega).$$

Da auch $\varphi(a)$ meßbar ist, und in seinem Definitionsbereiche, dem m-Raume, $0 \leqq a \leqq Max(f(x))$ das gewöhnliche Lebesguesche Maß als l-Maß angesehen werden kann, können wir auch seine Maßfunktion $\psi(b)$ bilden. $\psi(b)$ ist (vgl. w. o.) in $0 \leqq b \leqq \mu(\Omega)$ definiert, monoton nichtfallend und nach rechts halbstetig, und es ist stets $0 \leqq \psi(b) \leqq Max(f(x))$. Übrigens ist $\psi(b)$ die obere Grenze aller a mit $\varphi(a) \leqq b$, also in einem gewissen Sinne die

[21] Annals of Math., 3212 (1931), S. 194, Satz 1.

[22] Offenbar gilt dies dann auch für die Summe aller $\xi \leqq b - \frac{1}{n}$-Intervalle ($n = 1, 2, \cdots$), d. i. $\xi < b$, und für die Differenzmenge $a < \xi < b$. Hieraus folgert man es mühelos nacheinander für alle offenen, abgeschlossenen, und Borelschen ξ-Mengen.

Inverse von $\varphi(a)$; *dasselbe gilt, wenn* $\varphi(a)$ *und* $\psi(b)$ *vertauscht werden (vgl. a. a. O., S. 193 unten).*

Der zu beweisende Satz aber lautet so:

SATZ 3. *Seien* Ω, μ^* *wie in Def.* 6; f, φ, ψ *ebenfalls wie dort, d. h.* $f(x)$ *in* Ω *definiert und mit reellen Zahlen als Werten, nicht negativ und beschränkt,* $\varphi(a)$ *seine Maßfunktion,* $\psi(b)$ *diejenige von* $\varphi(a)$.

Sei $g(u)$ *für reelle Zahlen definiert und mit solchen als Werten, aber sonst ganz beliebig*[23]. $g(f(x))$ *ist dann und nur dann meßbar (vgl. Def.* 6), *wenn* $g(\psi(b))$ *es ist (dieses im gewöhnlichen Lebesgueschen Sinne), und es gilt dann*

$$\int_{\Omega} g(f(x))\, dv_x = \int_0^{\mu(\Omega)} g(\psi(b))\, db$$

(d. h. wenn eine Seite Sinn hat, dann auch die andere, und sie sind einander gleich).

Dieser Satz ist in zwei Beziehungen allgemeiner, als der in Anm. [21] angeführte: Erstens braucht Ω nicht, wie dort, eine Zahlenmenge zu sein, und μ^* das gewöhnliche Lebesguesche äußere Maß[24]; zweitens waren dort alle Schlußweisen nur in einer Richtung (von $g(\psi(b))$ auf $g(f(x))$) begründet, während die in der anderen Richtung fehlten[25]. Wir gehen nun zu seinem Beweise über.

Es genügt offenbar, zu zeigen: wenn von den zwei Mengen $g(f(x)) \leqq c$ und $g(\psi(b)) \leqq c$ (c fest) die eine meßbar ist, so ist es auch die andere, und beide haben dasselbe Maß. Oder wenn Ξ die u-Menge $g(u) \leqq c$ ist: wenn vom $u = f(x)$-Urbild und vom $u = \psi(b)$-Urbild von Ξ das eine meßbar ist, so ist es auch das andere, und beide haben dasselbe Maß. Wir beweisen dies für alle Ξ.

Betrachten wir das Zahlenintervall $0 \leqq u \leqq \mathrm{Max}(f(x))$. Es ist ein m-Raum, Ω_0. Wir definieren in ihm zwei l-Maße: erstens μ_1^* durch $\mu_1^*(\Xi) = \mu^*$ des f-Urbildes von Ξ, zweitens μ_2^* durch $\mu_2^*(\Xi) =$ gewöhnliches Lebesguesches äußeres Maß des ψ-Urbildes von Ξ. Die zu beweisende Behauptung lautet dann: μ_1^* ist mit μ_2^* identisch, oder: $\varphi_0(x) = x$ ist eine maßtreue Abbildung des Ω_0 mit μ_1^* auf das Ω_0 mit μ_2^* (vgl. Def. 4). Nach Satz 2 genügt es hierzu, zu zeigen, daß die zweite Bedingung von

[23] Es braucht z. B. keineswegs meßbar zu sein!

[24] Bei den Anwendungen zum Beweise des Ergodensatzes (vgl. den nächstfolgenden Artikel des Verfassers sowie den in Proc. Nat. Ac. 18, January 1931) ist gerade diese Allgemeinheit wesentlich. Am zweit-a. O. wurde dies mittels der maßtreuen Abbildbarkeit der auftretenden Mengen auf reelle Zahlenmengen umgangen. Nun kann zwar das Bestehen dieser Abbildbarkeit, nach einer mündlichen Mitteilung von Herrn M. H. Stone an den Verfasser, allgemein bewiesen werden — es ist aber wohl von Interesse, auch einen allgemeinen direkten Beweis zu besitzen, besonders da er ebenso verläuft wie der Beweis des Spezialfalles.

[25] Dies reicht zwar für die Anwendungen aus, ist aber prinzipiell unbefriedigend.

Def. 4 erfüllt ist, welche im vorliegenden Falle dies besagt: wenn Ξ μ_2^*-meßbar ist, so ist es auch μ_1^*-meßbar, und die beiden Maße sind einander gleich. Für welche Ξ trifft dies nun wirklich zu?

Wenn Ξ das Intervall $u \leqq a$ ist (d. h. der in $0 \leqq u \leqq \mathrm{Max}\,(f(x))$ gelegene Teil desselben), so ist das f-Urbild die Menge $f(x) \leqq a$, also meßbar, und vom Maße $\varphi(a)$, daher ist Ξ μ_1^*-meßbar und $\mu_1^*(\Xi) = \varphi(a)$. Das ψ-Urbild dagegen ist die Menge $\psi(b) \leqq a$, d. h. $b <$ oder $\leqq \varphi(a)$ (natürlich $b \geqq 0$), also meßbar und vom Maße $\varphi(a)$, daher ist Ξ auch μ_2^*-meßbar, und $\mu_2^*(\Xi) = \varphi(a)$. Das Intervall $u \leqq a$ ist also ein solches Ξ. Also auch die Summe der Intervalle $u \leqq b - \frac{1}{n}$ $(n = 1, 2, \cdots)$, d. i. $u < b$; also auch die Differenzmenge $a < u < b$ (vgl. Anm. [22]). Und hieraus folgt dasselbe nacheinander für alle offenen, abgeschlossenen und Borelschen u-Mengen (vgl. Anm. [22]). Ist $\mu_2^*(\Xi) = 0$, so ist, wie wir schon oft schlossen, $\Xi \subset H$, H Borelsch, $\mu_2(H) = 0$. Nach dem soeben Gesagten ist also auch $\mu_1(H) = 0$, also auch Ξ μ_1-meßbar und $\mu_1(\Xi) = 0$. Also: auch die Ξ mit $\mu_2^*(\Xi) = 0$ haben die obige Eigenschaft. Und da, wie wir schon oft schlossen, jedes μ_2^*-meßbare Ξ Summe eines Borelschen und eines mit dem μ_2^*-Maß 0 ist, haben es auch diese. D. h.: die genannte Eigenschaft liegt bei allen Ξ, die für sie überhaupt in Frage kommen, vor. Damit ist der Beweis durchgeführt.

ZUR OPERATORENMETHODE IN DER KLASSISCHEN MECHANIK[1].

VON J. v. NEUMANN, PRINCETON.

Einleitung.

1. Die von Herrn B. O. Koopman entdeckte Operatorenbehandlung der klassischen Mechanik[2] hat in der kurzen Zeit seit ihrem Bekanntwerden eine sehr beachtenswerte Durchschlagskraft an bis dahin unbezwungenen Problemen dieser Disziplin bewiesen[3]. Die Hoffnung, daß sie in der Zukunft Wesentliches zum Fortschritt derselben beitragen wird, erscheint daher als berechtigt. Ihr weiterer Ausbau, der in erster Linie eine nähere Analyse der durch sie eingeführten Operatoren erheischt, ist darum höchst erwünscht. Einen Ansatz zu diesem Programm stellt die vorliegende Abhandlung dar, indem sie die hauptsächlichsten Zusammenhänge zwischen mechanischen Fragen und operatorentheoretischen Problemen angibt und einige der letzteren löst, für andere aber die gegenwärtig wahrscheinlichsten Vermutungen aufstellt.

Wir verwenden die Bezeichnungen von Koopman (vgl. a. a. O. Anm. [2], [3]): Φ sei der Phasenraum eines mechanischen Systems, P sein allgemeiner Punkt, S_t: $P \to P_t = S_t P$ die durch die den mechanischen Differentialgleichungen gemäß in der Zeit t erfolgte Bewegung beschriebene „Strömung" in Φ. Nach dem Liouvilleschen Satze ist S_t maßtreu: d. h. jede (im Lebesgueschen Sinne) meßbare Teilmenge von Φ wird durch S_t auf eine ebensolche gleichen Maßes abgebildet[4] und umgekehrt[5]. Sei Ω eine gegenüber allen S_t invariante Teilmenge von Φ, d. h. eine „Integralfläche" der mechanischen Bewegung, man führt dann auf dieser Fläche Ω in bekannter

[1] Received April 10, 1932.

[2] Proc. Nat. Ac. 17, Mai 1931, S. 315.

[3] Vgl. die Abhandlung des Verfassers Proc. Nat. Ac. 18, Jan. 1932, S. 70 (Beweis des Ergodensatzes) sowie B. O. Koopman und der Verfasser, Proc. Nat. Ac. 18, March 1932, S. 255, (Beweis des „Verrührungssatzes"). Vgl. ferner E. Hopf, Proc. Nat. Ac. 18, Jan. 1932, S. 93, und Febr. 1932, S. 204.

[4] In der Mechanik wird dies in der Regel nur für Parallelepipede bewiesen. Von diesen kann man es aber auf die Summen endlich vieler Parallelepipede übertragen sowie auf die Limites aufsteigender Folgen solcher Mengen, d. i. auf alle offenen Mengen. Sind $O_1, O_2, \cdots$ offen, so gilt es auch für die Durchschnitte O_1, $O_1 O_2$, $O_1 O_2 O_3, \cdots$ (diese sind offen), also auch für ihren Limes $O_1 O_2 \cdots$. Da jede Menge vom Maße 0 Teil eines $O_1 O_2 \cdots$ vom Maße 0 ist, gilt es auch für diese. Und da jede meßbare Menge Differenz eines $O_1 O_2 \cdots$ und einer Menge vom Maße 0 ist, gilt es allgemein.

[5] Die Umkehrung folgt aus der entsprechenden Eigenschaft der inversen Abbildung S_{-t}, oder auch direkt nach Satz 2 der vorhergehenden Abhandlung des Verfassers.

Weise ein Volumelement $d\omega$ und eine Dichte $\varrho = \varrho(P) > 0$ ein, so daß das Gewicht $\varrho\, d\omega$ S_t-invariant ist: d. h. wenn für die Teilmengen $\mathcal{A}$ von Ω ein Lebesguesches Maß durch $\mu(\mathcal{A}) = \int_{\mathcal{A}} \varrho\, d\omega$ definiert wird, so ist S_t auch in Ω maßtreu[6].

Die S_t bilden in φ (wie in Ω) eine einparametrige Gruppe, d. h. es ist stets: $S_t(S_s P) = S_{t+s} P$, symbolisch: $S_t S_s = S_{t+s}$. Sie sind, wie wir sahen, eineindeutige und maßtreue Abbildungen von Φ (und von Ω) auf sich selbst. Außerdem sind sie in ihrer Abhängigkeit von P, t stetig, ja sogar stetig differentiierbar.

Diese letztere Eigenschaft werden wir im folgenden nie benutzen, sie ist für die Koopmansche Operatorenmethode unwesentlich. Wir wollen daher das allgemeine Problem, das über die mechanischen „Strömungen" hinausgeht, gleich formulieren. Den Phasenraum φ, bzw. die Integralflächen Ω, verallgemeinern wir so:

D_1. *Sei Ω ein beliebiger m-Raum, d. h. metrisch separabel, und vollständig*[7], *für dessen Teilmengen $\mathcal{A}$ ein l-Maß $\mu^*(\mathcal{A})$ (d. h. eines mit den Eigenschaften des Lebesgueschen äußeren Maßes) definiert ist*[8].

Die reellen Zahlen t, mit der gewöhnlichen für sie gültigen Topologie, bilden ebenfalls einen m-Raum T mit dem gewöhnlichen Lebesgueschen äußeren Maß als l-Maß, $\tilde{\mu}^*(\tilde{\mathcal{A}})$.

Der „Produktraum" $\Omega \times T$ aller Paare $[P, t]$[9] *ist ebenfalls ein m-Raum, und ein l-Maß kann in ihm so definiert werden:*

α) Seien O in Ω und $\tilde{O}$ in T offene Mengen, $O \times \tilde{O}$ die Menge der $[P, t]$, P in O, t in $\tilde{O}$. Dann setzen wir

$$\Im(O \times \tilde{O}) = \mu(O)\, \tilde{\mu}(\tilde{O}).$$

β) Sei $\overline{\mathcal{A}}$ eine Menge in $\Omega \times T$. Betrachten wir alle Folgen $O_1, \tilde{O}_1, O_2, \tilde{O}_2, \cdots$ mit $\overline{\mathcal{A}} \subseteqq O_1 \times \tilde{O}_1 + O_2 \times \tilde{O}_2 + \cdots$, die untere Grenze ihrer $\Im(O_1 \times \tilde{O}_1) + \Im(O_2 \times \tilde{O}_2) + \cdots$ heiße $\overline{\mu}^*(\overline{\mathcal{A}})$.

Diese Bildung ist dem mehrdimensionalen Lebesgueschen äußeren Maße analog, und sein l-Maßcharakter wird entsprechend bewiesen[10].

Der Sinn von Begriffen wie „meßbare Funktion", „Integral" usw. in Ω, T, $\Omega \times T$ ist nunmehr klar.

---

[6] Die allgemein übliche Diskussion beschränkt sich auch hier auf Parallelepipede, was dann nach Anm. [4], [5] verallgemeinert werden kann.

[7] Vgl. Def. 1 der vorhergehenden Abhandlung des Verfassers.

[8] Vgl. Def. 2 a. a. O. Anm. [7].

[9] Vgl. z. B. Hausdorff, „Mengenlehre", Berlin 1927, S. 102. Ist $\overline{P, Q}$ die Distanz in Ω, so ist die in $\Omega \times T$ etwa $\overline{[P, t], [Q, s]} = \sqrt{(\overline{P, Q})^2 + (t-s)^2}$.

[10] Vgl. z. B. Carathéodory, „Reelle Funktionen", Berlin 1918, S. 229—237 sowie 274 ff. Die weiteren Eigenschaften des mehrdimensionalen Maßes übertragen sich gleichfalls ohne weiteres auf unseren Fall, insbesondere der Satz von Fubini, S. 621—625.

Das Analogon der „Strömung“ S_t aber ist dies:

D_2. *Für jedes t sei eine eineindeutige maßtreue Abbildung S_t: $P \to P_t = S_t P$ von Ω auf sich selbst gegeben. Die S_t mögen eine einparametrige Gruppe bilden, d. h. es sei stets: $S_t(S_s P) = S_{t+s} P$, symbolisch $S_t S_s = S_{t+s}$. Als P, t-Funktion sei $S_t P$ meßbar*[11].

Es erweist sich u. U. als zweckmäßig, noch etwas weniger zu verlangen:

D_2'. *Mengen vom Maße 0 (wir nennen sie N_t', N_t'', $N_{t,s}$) werden in D_2 als Ausnahmen zugelassen: d. h. S_t braucht nur $\Omega - N_t'$ auf $\Omega - N_t''$ eineindeutig und maßtreu abzubilden (in N_t' sei S_t undefiniert); $S_t(S_s P) = S_{t+s} P$ braucht nur außerhalb von $N_{t,s}$ zu gelten*[12]. *Letzteres heiße symbolisch $S_t S_s \cong S_{t+s}$.*

Eine Schar S_t nach D_2' nennen wir eine „allgemeine Strömung“. Gilt sogar D_2 und ist $S_t P$ in P, t stetig oder stetig differentiierbar usw., so heiße die Schar S_t bzw. eine „stetige Strömung“ eine „stetig-differentiierbare Strömung“ usw. Entspringt sie schließlich, wie es eingangs der Fall war, einem mechanischen Problem, so heiße sie eine „mechanische Strömung“. — Zwei allgemeine Strömungen gelten als äquivalent, wenn sich ihre S_t für jedes gegebene t nur in einer P-Menge vom Maße 0 unterscheiden.

Eine weitere wichtige Begriffsbildung ist diese:

D_3. *Ω', Ω'' seien m-Räume mit l-Maßen ($\mu_1'^*$ bzw. $\mu_2''^*$) nach D_1, $\Sigma: P' \to P'' = \Sigma P'$ sei eine eineindeutige und maßtreue Abbildung von Ω' minus eine Menge vom Maße 0 auf Ω'' minus eine Menge vom Maße 0. Jeder allgemeinen Strömung S_t' in Ω' ordnet dieses Σ eine allgemeine Strömung $S_t'' = \Sigma S_t' \Sigma^{-1}$ in Ω'' zu und umgekehrt.*

Zwei allgemeine Strömungen S_t', S_t'' (in Ω' bzw. Ω'') sollen isomorph heißen, wenn ein solches Σ existiert, das sie (bis auf eine Menge vom Maße 0) ineinander überführt: $S_t'' \cong \Sigma S_t' \Sigma^{-1}$.

Isomorphe Strömungen können wir als nicht wesentlich verschieden ansehen, da ihr einziger Unterschied darin liegt, daß die Punkte P' von Ω' durch die entsprechenden $P'' = \Sigma P'$ von Ω'' ersetzt wurden. Bei solchen Problemen aber, bei denen es nur auf die Volumina der auftretenden Mengen ankommt (wir werden weiter unten auseinandersetzen, welche Probleme wir meinen), ist dies wegen der Maßtreue von Σ unwesentlich.

Die Stetigkeit, oder gar stetige Differentiierbarkeit einer Strömung kann bei Ersetzung durch eine isomorphe verlorengehen — dies sind keine iso-

[11] D. h.: wenn O eine offene Menge in Ω ist, so sei die durch $S_t P$ in O bestimmte $[P, t]$-Menge in $\Omega \times T$ meßbar.

[12] Da N_t', N_t'', N_{t+s} von t, s abhängen und diese über unabzählbar viele Werte laufen, können wir die genannten Ausnahmemengen nicht zu einer Gesamtmenge vom Maße 0 zusammenfassen.

morphieinvarianten Eigenschaften. Vermutlich kann sogar zu jeder allgemeinen Strömung eine isomorphe stetige Strömung gefunden werden[13], vielleicht sogar eine stetig-differentiierbare, oder gar eine mechanische. Dies mag es rechtfertigen, daß hier an Stelle der eigentlich interessanten mechanischen Strömungen alle allgemeinen untersucht werden.

2. Die mechanischen Probleme, die mit der Koopmanschen Methode angreifbar sind, sind vom folgenden Typus: Welche geometrische Wahrscheinlichkeit im Phasenraume haben gewisse, mit Hilfe der mechanischen Bewegung beschriebene Ereignisse — d. h. welches Maß haben gewisse, mit Hilfe der S_t definierte Teilmengen von Ω? Es handelt sich also um isomorphieinvariante Eigenschaften der S_t. Einige Beispiele[14]:

a) Sei Λ gegeben, $Z(\Lambda, P, t, s) = \tilde{\mu}$ (τ-Menge der $S_\tau P$ in Λ, $t \leqq \tau < s$). Gilt $\dfrac{Z(\Lambda, P, t, s)}{s-t} \to \dfrac{\mu(\Lambda)}{\mu(\Omega)}$ bei $s-t \to +\infty$ für alle P mit Ausnahme einer Menge vom Maße 0?

b) Gilt $\dfrac{Z(\Lambda, P, t, s)}{s-t} \to \dfrac{\mu(\Lambda)}{\mu(\Omega)}$ bei $s-t \to +\infty$ wenigstens im Sinne der Mittelkonvergenz einer P-Funktion[15]?

c) Seien Λ, M gegeben, $Z'(\Lambda, M, t, s) = \int_t^s \mu(S_t \Lambda \cdot M)\, d\tau$. Gilt $\dfrac{Z'(\Lambda, M, t, s)}{s-t} \to \dfrac{\mu(\Lambda)\mu(M)}{\mu(\Omega)}$ bei $s-t \to +\infty$?

d) Gilt sogar $\mu(S_t \Lambda \cdot M) \to \dfrac{\mu(\Lambda)\mu(M)}{\mu(\Omega)}$ bei $t \to \pm\infty$?

e) Gilt wenigstens $\mu(S_t \Lambda \cdot M) \to \dfrac{\mu(\Lambda)\mu(M)}{\mu(\Omega)}$ bei $t \to \pm\infty$, wenn eine feste t-Menge I von der Dichte 0[16] ausgeschlossen wird?

a), b), c) sind verschiedene Fassungen der sog. Quasi-Ergoden-Hypothese[17], u. zw. von abnehmender Schärfe: aus a) folgt b), aus b), wie wir sehen

---

[13] Der Verfasser hofft, hierfür demnächst einen Beweis anzugeben.

[14] Mit Λ, M werden wir meßbare Teilmengen von Ω mit endlichem Maße bezeichnen, $S_t \Lambda$ ist das durch S_t vermittelte Bild von Λ, $\Lambda \cdot M$ der Durchschnitt von Λ und M.

[15] D. h. ist $\int_\Omega \left(\dfrac{Z(\Lambda, P, t, s)}{s-t} - \dfrac{\mu(\Lambda)}{\mu(\Omega)} \right)^2 d\nu_P \to 0$ für $s-t \to +\infty$. Mit $\int_\Omega \cdots d\nu_P$ bezeichnen wir das auf unser μ gegründete Integral. Natürlich ist auch dies eine Aussage über Maße.

[16] Dichte 0 bedeutet: $\dfrac{\tilde{\mu}(\tau\text{-Menge der } \tau \text{ in } I,\ -t \leqq \tau \leqq t)}{2t} \to 0$ für $t \to +\infty$. In den Fällen, die uns interessieren, ist I als Vereinigung einer sich im Unendlichen häufenden Intervallfolge wählbar.

[17] Dieselbe ist in der physikalischen Literatur, trotz ihrer Wichtigkeit, mathematisch recht ungenau formuliert. Vgl. dazu die Abhandlungen des Verfassers Proc. Nat. Ac. 18, Jan. 1932, S. 70 und 18, March 1932, S. 263.

werden, c). Für die physikalischen Anwendungen der Quasi-Ergoden-Hypothese[18] ist, wie eine nähere Betrachtung lehrt, b) die notwendige und hinreichende Fassung[18]. d), e) sind Verschärfungen von c), also Verschärfungen der Quasi-Ergoden-Hypothese (für die Fassungen a), b) kommt offenbar etwas Derartiges nicht in Frage), u. zw. folgt offenbar e) aus d). Ihr Sinn ist dieser: jede Teilmenge $\mathfrak{A}$ von Ω wird durch die Strömung S_t derartig verschmiert, daß sie für große t in jedes $\mathfrak{M}$ eindringt, und Ω homogen erfüllt $\left(\frac{\mu(S_t\mathfrak{A}\cdot\mathfrak{M})}{\mu(\mathfrak{A})} \to \frac{\mu(\mathfrak{M})}{\mu(\Omega)}\right.$, d. h. in $\mathfrak{M}$ eingedrungener relativer Anteil von $\mathfrak{A} \to$ relativer Anteil von $\mathfrak{M}$ an $\left.\Omega\right)$; oder: die Wahrscheinlichkeit dafür, daß P zur Zeit $\tau = 0$ in $\mathfrak{A}$ und zur Zeit $\tau = t$ in $\mathfrak{M}$ liegt $\left(\frac{\mu(S_t\mathfrak{A}\cdot\mathfrak{M})}{\mu(\Omega)}\right.$, P bewegt sich mit der Strömung $\left.S_\tau!\right)$ strebt gegen die Wahrscheinlichkeit dessen, daß zwei unabhängige Punkte P, Q in $\mathfrak{A}$ bzw. $\mathfrak{M}$ liegen $\left(\frac{\mu(\mathfrak{A})\,\mu(\mathfrak{M})}{(\mu(\Omega))^2}\right)$; d. h.: das Wahrscheinlichkeitsverhalten von P, S_tP strebt gegen die statistische Unabhängigkeit.

Die notwendigen und hinreichenden Bedingungen für die Gültigkeit von b), von c) (es sind dieselben), sowie von e), und einiges über d), konnten mit Hilfe der Koopmanschen Methode aufgefunden werden[19].

Eine naheliegende Frage ist, warum wir in a)—e) alle meßbaren $\mathfrak{A}$, $\mathfrak{M}$ zulassen, und nicht nur solche, die bei Beantwortung physikalischer Fragen tatsächlich in Frage kommen, d. h. für welche die Zugehörigkeit von P durch Messungen beschränkter Genauigkeit mit hoher Wahrscheinlichkeit entschieden werden kann. Also etwa abgeschlossene oder offene $\mathfrak{A}$, oder gar nur Parallelepipede. Der Grund hierfür ist ein zweifacher:

α) Entgegen der allgemeinen Auffassung ist für jedes meßbare $\mathfrak{A}$ die Zugehörigkeit von P im obigen Sinne „physikalisch" entscheidbar.

β) Wenn eines der a)—e) für alle Parallelepipede gilt, so gilt es für alle meßbaren $\mathfrak{A}$ von endlichem Maß.

Beides wurde a. a. O. Anm. [3] gezeigt, wir werden es weiter unten erneut beweisen. α), β) sind ein weiterer Beleg dafür, daß die wichtigsten physikalischen Eigenschaften isomorphieinvariant sind (ohne Rücksicht auf Stetigkeitsfragen!)

---

[18] D. i. die Zurückführung der kinetischen Gastheorie auf die Maxwell-Boltzmannsche statistische Mechanik, bzw. die Gibbssche „Mikrokanonische Gesamtheit". Vgl. a. a. O. Anm. [17].

[19] Siehe a. a. O. Anm. [3]. Unmittelbar nachher bewies Herr G. D. Birkhoff mit einer anderen Methode, daß unter denselben Bedingungen sogar a) gilt, Proc. Nat. Ac. 17, Dec. 1931, S. 650. Vgl. auch die historische Darstellung von B. O. Koopman und G. D. Birkhoff, Proc. Nat. Ac. 18, March 1932, S. 279.

3. Der Grundgedanke der Koopmanschen Methode ist dieser: Die meßbaren (komplexwertigen) Funktionen $f(P)$ in Ω mit endlichem $\int_\Omega |f(P)|^2 dv_P$ bilden einen Hilbertschen Raum $\mathfrak{H}$, falls das innere Produkt (f, g) durch $\int_\Omega f(P)\overline{g(P)}\,dv_P$ definiert wird (also der Betrag $\|f\| = \sqrt{(f,f)}$ durch $\sqrt{\int_\Omega |f(P)|^2\,dv_P}$)[20]. Der Funktionaloperator $U_t: f(P) \to f(S_t P) = U_t f(P)$ ist linear, und wegen D_2' (Maßtreue von S_t) unitär[21]. Wegen $S_t S_s \cong S_{t+s}$ ist $U_t U_s = U_{t+s}$. Und aus der (in D_2' übernommenen) letzten Forderung von D_2 folgt (vgl. § III 2): $(U_t f, g) = \int_\Omega f(S_t P)\overline{g(P)}\,dv_P$ ist eine meßbare Funktion von t. Also:

D_4. *Die U_t sind unitäre Operatoren, die eine einparametrige Gruppe bilden: $U_t U_s = U_{t+s}$. U_t hängt meßbar von t ab, d. h. jedes $(U_t f, g)$ ist eine meßbare Funktion von t.*[22]

Wenn U_t stetig von t abhängt, ist ein Satz von Herrn M. H. Stone anwendbar[23], wonach ein hypermaximaler Operator A mit der Zerlegung der Einheit $E(\lambda)$ existiert[24], so daß

$$(U_t f, g) = \int e^{it\lambda}\, d(E(\lambda) f, g)$$

ist (Stieltjessches Integral!), oder symbolisch:

$$U_t = e^{itA} = \int e^{it\lambda}\, dE(\lambda).$$

Wie der Verfasser a. a. O. Anm. [22] bewies, ist die Stetigkeit nicht erforderlich, es genügt vielmehr hierzu die durch D_4 gewährleistete meßbare Abhängigkeit von t: aus dieser folgt die Stetigkeit und die Gültigkeit des Stoneschen Satzes (vgl. Satz 1, 2, a. a. O.).

Die Aussagen b)—e) können nun alle mit Hilfe der U_t ausgedrückt werden, und ihre weitere Diskussion erfolgt dann auf Grund der Stoneschen Darstellung der U_t: $U_t = e^{itA}$ (vgl. a. a. O. Anm. [2], [3] sowie im weiteren Verlauf dieser Abhandlung). Bei den bisherigen Untersuchungen war es dabei so, daß nur die in D_4 angegebenen Eigenschaften der U_t, die der

---

[20] Die Abhandlungen des Verfassers Math. Ann. 102/1, 1929, S. 49, und Math. Ann. 102/3, 1929, S. 370, werden mehrfach zitiert werden, u. zw. als E bzw. A. Zum Obigen vgl. etwa E, S. 108—111 (Anhang I), es handelt sich eigentlich um den Fischer-Rieszschen Satz.

[21] Es ist $\int_\Omega |f(P)|^2\,dv_P = \int_\Omega |f(S_t P)|^2\,dv_P$, d. h. $\|f\| = \|U_t f\|$.

[22] Vgl. die zweit-vorhergehende Abhandlung des Verfassers in diesem Bande.

[23] Proc. Nat. Ac. 16, Febr. 1930, S. 172, sowie die nächstfolgende Abhandlung von Herrn M. H. Stone in diesem Bande.

[24] Vgl. E S. 72, 91, 92.

Stoneschen Darstellung gleichwertig sind, eine Rolle spielten, nur bei d) kommt es noch auf eine feinere Eigenschaft des $E(\lambda)$-Spektrums an[25]. Bei tiefergehenden Problemen ist aber zu erwarten, daß die spezielle Entstehungsweise der U_t (d. i. $U_t f(P) = f(S_t P)$) eine Rolle spielen wird. Dieselbe hat nämlich weitere, in D_4 nicht erwähnte Eigenschaften der U_t zur Folge, wie sie a. a. O. Anm. [2] von Koopman angegeben wurden. Eine von ihnen ist

D_5. $$U_t f \cdot U_t g = U_t (f \cdot g),$$

falls $f, g, f \cdot g$ alle zum Hilbertschen Raume gehören (der Beweis ist klar). Diese Eigenschaften haben nun die Konsequenz, daß nicht jede Schar $E(\lambda)$ bzw. nicht jedes A auftreten kann, was für die Probleme von unserem Typus von Belang ist. Insbesondere werden die Möglichkeiten für die Struktur des $E(\lambda)$-Spektrums eingeschränkt.

Ein weiterer Umstand, der Aufmerksamkeit verdient, ist dieser. Sind Ω', Ω'' im Sinne von D_4 vermittels der Abbildung Σ isomorph und $S_t', S_t'' = \Sigma S_t' \Sigma^{-1}$ einander entsprechende allgemeine Strömungen, so gilt für ihre Operatorenscharen U_t', U_t'' analog $U_t'' = V U_t' V^{-1}$, falls der unitäre Operator V durch $f(P') \to f(\Sigma^{-1} P'') = V f(P'')$ definiert wird. Daraus folgt (da V unsere Hilbertschen Räume isomorph aufeinander abbildet) $E'(\lambda) = V E''(\lambda) V^{-1}$, $A' = V A'' V^{-1}$. D. h. alle spektralen Eigenschaften stimmen überein, wie es wegen ihrer offenbaren Isomorphieinvarianz auch nicht anders zu erwarten war. Die spektralen Eigenschaften sind aber auch dann invariant, wenn wir bloß die Transformation $U_t'' = V U_t' V^{-1}$, mit unitärem V haben, ohne daß die Existenz eines Σ, sowie der obige Zusammenhang zwischen ihm und V, gefordert würde. Ja alle Eigenschaften, die mit Hilfe der U_t formuliert werden können, sind so, wir wollen sie daher „unitär-invariant" nennen. So sind b)—e) unitär-invariant, a) dagegen zunächst nur isomorphieinvariant. Daß die Koopmansche Methode wirklich alle Wahrscheinlichkeitsfragen der klassischen Mechanik erfaßt, wäre belegt, wenn wir wüßten, daß alle isomorphieinvarianten Eigenschaften auch unitär-invariant sind. D. h. wenn wir jedesmal, wenn zwei Ω', Ω'' mit ihren S_t', S_t'' (nach D_1, D_2') gegeben sind, und für ihre U_t', U_t'' $U_t'' = V U_t' V^{-1}$ mit unitärem V gilt, eine eineindeutige maßtreue Abbildung Σ von Ω' auf Ω'' finden könnten, für die $S_t'' = \Sigma S_t' \Sigma^{-1}$ gilt.[26]

---

[25] Nämlich auf den sog. Hellingerschen Typus seines Streckenspektrums. Vgl. die zweite in Anm. [3] genannte Abhandlung. Über den Hellingerschen Typus vgl. E. Hellinger, Dissertation, Göttingen, 1907, sowie H. Hahn, Monatshefte, 23, 1912, S. 161.

[26] D. h. für $W: f(P') \to f(\Sigma^{-1} P'') = W f(P'')$ soll $U_t'' = W U_t' W^{-1}$ sein. $W = V$ wird natürlich nicht verlangt.

4. Die Einteilung dieser Abhandlung ist diese: In § I diskutieren wir, wie die Zugehörigkeit zu irgendeinem meßbaren A in Ω physikalisch entscheidbar ist. In § II werden die Gültigkeitsbedingungen der Quasi-Ergoden-Hypothese diskutiert und gezeigt, wie jede allgemeine Strömung als Vereinigung einer Schar solcher Strömungen dargestellt werden kann, deren jede die Q. E. H. erfüllt — oder, wie es abkürzend heißen soll, ergodisch ist.[27] In den folgenden Paragraphen betrachten wir demgemäß nur ergodische Strömungen. § III. handelt von den charakteristischen Eigenschaften der Koopmanschen U_t-Scharen, insbesondere wird gezeigt, daß D_5 charakteristisch ist. In § IV. werden für den Fall, daß U_t ein „reines Punktspektrum" hat, diejenigen Spektra angegeben, die bei Koopmanschen U_t auftreten können.[28] Ferner wird gezeigt, daß in diesem Falle das Spektrum allein die U_t bis auf eine Isomorphie (Σ, vgl. 3.) festlegt, d. h. daß hier die am Ende von 3. formulierte Vermutung gilt. § V. ist den Fällen mit Streckenspektrum gewidmet. Hier sind analoge Sätze zu vermuten, bewiesen kann aber vorläufig nur weniges werden. Trotzdem ist gerade der Fall eines reinen Streckenspektrums der eigentlich interessante. Denn einerseits gilt dann und nur dann e) (vgl. das dazu in 3. Gesagte, dieser Satz wurde in der zweiten der in Anm. [3] genannten Abhandlungen bewiesen), d. h. solche Strömungen haben keine über lange Zeiträume gültigen Eigenschaften (vgl. das in § V hierüber zu Sagende); und andererseits ist zu vermuten, daß Strömungen, die nicht ein reines Streckenspektrum haben, seltene Ausnahmen sind. Diese letztere Vermutung wird in § VI durch Diskussion einer ausgedehnten Klasse zweidimensionaler Flüsse belegt, da diese fast alle solche Spektra haben.

Zum Schluß sei noch auf die interessante Analogie zwischen Koopmans Operatoren $U_t = e^{itA}$ und den Operatoren der Quantenmechanik hingewiesen. Die Schrödingersche Wellenfunktion φ (definiert im Zustandsraume des mechanischen Systems und nicht wie unsere f im Phasenraume!) gehorcht bekanntlich in ihrer Abhängigkeit vom Zeitparameter t der Differentialgleichung $\frac{h}{2\pi i}\frac{\partial}{\partial t}\varphi = H\varphi$. Hier ist h das Plancksche Wirkungsquantum und H der Energieoperator. Hieraus folgt sofort $\varphi = e^{it\cdot\frac{2\pi}{h}H}\varphi_{(t=0)}$[29], so daß hier die unitären Operatoren $\hat{U}_t = e^{it\cdot\frac{2\pi}{h}H}$ eine fundamentale Rolle spielen. Die Analogie, die durch das Nebeneinanderstellen von A und

[27] § I und der Hauptteil von § II wurden bereits in den Proc. Nat. Ac. 18, Jan. 1932, S. 70, publiziert.

[28] Bei all diesen Bedingungen wurde der notwendige Charakter von Koopman, a. a. O. Anm. [2], bewiesen. Neu ist der wesentlich schwerere Nachweis ihrer Hinreichendheit.

[29] Meistens wird diese Beziehung mit $-H$ statt H geschrieben.

$\frac{2\pi}{h} H$ entsteht, ist auffallend[30], und es ist möglich, sie zum Nachweis des stetigen Übergehens der Quantenmechanik in die klassische (für $h \to 0$) auszubauen. Trotzdem scheinen wesentliche mathematische Unterschiede zwischen diesen Operatorenscharen zu bestehen. Denn für ein mechanisches System, das in ein endliches Volumen eingesperrt ist, scheint in der Quantenmechanik stets ein reines Punktspektrum vorzuliegen[31], während im klassisch-mechanischen Problem das reine Streckenspektrum der allgemeine Fall zu sein scheint (vgl. § VI).

I. Meßbare Mengen und physikalische Messungen.

1. Ω sei ein Raum nach D_1, sein Gesamtmaß endlich: $\mu(\Omega) > 0$, $< +\infty$. Seine Punkte P mögen die Zustände eines mechanischen Systems repräsentieren (Ω sei etwa ein Phasenraum oder Ähnliches). Jede Eigenschaft dieses Systems wird dann durch eine geeignete Teilmenge $\mathfrak{A}$ von Ω repräsentiert, die aus den Bildpunkten P aller Zustände besteht, die diese Eigenschaft besitzen. Dem l-Maß $\mu(\mathfrak{A})$ in Ω schreiben wir die Bedeutung zu, daß wir für die Gültigkeit einer solchen Eigenschaft, d. h. für die Zugehörigkeit von P zu $\mathfrak{A}$ die a priori Wahrscheinlichkeit $\frac{\mu(\mathfrak{A})}{\mu(\Omega)}$ ansetzen. Und wenn wir wissen, daß P einer gegebenen Menge $\mathfrak{M}$ angehört, $\mathfrak{A} \subseteq \mathfrak{M} \subseteq \Omega$, die Wahrscheinlichkeit $\frac{\mu(\mathfrak{A})}{\mu(\mathfrak{M})}$. (Die auftretenden $\mathfrak{A}$, $\mathfrak{M}$, $\mathfrak{N}$, $\cdots$ sollen alle meßbar sein.)

Eine physikalische Messung entscheidet immer, ob P einem gewissen $\mathfrak{A}$ angehört oder nicht. Die Tatsache, daß nur Messungen von beschränkter Genauigkeit möglich sind, daß sich aber die Genauigkeit beliebig steigern läßt, wird wohl am besten so beschrieben:

Z. Für jedes $n = 1, 2, \cdots$ ist Ω in eine endliche oder abzählbar-unendliche Anzahl von meßbaren Mengen $\mathfrak{N}_1^{(n)}, \mathfrak{N}_2^{(n)}, \cdots$ eingeteilt, von denen je zwei einen Durchschnitt vom Maß 0 haben, und deren Vereinigung Ω ist. Dabei sei die $n+1$-te Einteilung eine Unterteilung der n-ten, d. h. jedes $\mathfrak{N}_\nu^{(n)}$ sei genau die Vereinigung einiger $\mathfrak{N}_{\nu'}^{(n)}, \mathfrak{N}_{\nu''}^{(n)}, \cdots$. Ferner soll $\delta_n =$ Maximum der Diameter aller $\mathfrak{N}_\nu^{(n)}$ $(\nu = 1, 2, \cdots)$[32] für $n \to \infty$ gegen Null streben.

[30] Eine genauere Überlegung zeigt, daß sie vollkommener wird, wenn man die Differentialgleichung der Wellenfunktion durch diejenige des sog. statistischen Operators ersetzt (vgl. z. B. J. v. Neumann, Mathematische Grundlagen der Quantenmechanik, Berlin 1932, S. 186). Dieselbe ist aber ebenso gebaut und das weiter unten zu Sagende gilt auch für sie.

[31] Dies findet man z. B. bei van der Waerden, Die gruppentheoretische Methode in der Quantenmechanik, Berlin 1932, S. 5, recht klar formuliert.

[32] Der Diameter von $\mathfrak{A}$ ist das Maximum aller $\overline{x, y}$, x, y in $\mathfrak{A}$.

Unter einer Messung n-ter Genauigkeitsstufe verstehen wir eine, welche entscheidet, welchem der $N_1^{(n)}, N_2^{(n)}, \cdots$ P angehört[33].

Es sind Messungen aller Genauigkeitsstufen möglich.

Nehmen wir nun an, wir wollten entscheiden, ob P zu einer gegebenen Menge $\mathcal{A}$ gehört, und nehmen zu diesem Zweck eine Messung von der Genauigkeitsstufe n vor. Dann erfahren wir, zu welchem $N_1^{(n)}, N_2^{(n)}, \cdots$ P gehört, und zwar ist die a priori Wahrscheinlichkeit dafür, daß $N_\nu^{(n)}$ herauskommt, $\frac{\mu(N_\nu^{(n)})}{\mu(\Omega)}$. Und ist $N_\nu^{(n)}$ herausgekommen, so bedeutet die Zugehörigkeit zu $\mathcal{A}$ die zu $\mathcal{A} \cdot N_\nu^{(n)}$, $\mathcal{A} \cdot N_\nu^{(n)} \subseteq N_\nu^{(n)} \subseteq \Omega$. Die Wahrscheinlichkeit der Zugehörigkeit zu $\mathcal{A}$ ist also dann $\frac{\mu(\mathcal{A} \cdot N_\nu^{(n)})}{\mu(N_\nu^{(n)})}$. Sei nun ein $\varepsilon > 0$ gegeben. Ist $\frac{\mu(\mathcal{A} \cdot N_\nu^{(n)})}{\mu(N_\nu^{(n)})} > 1-\varepsilon$ oder $< \varepsilon$, so können wir auf die Frage „gehört P zu $\mathcal{A}$?“ antworten „ja“ bzw. „nein“, mit einer Wahrscheinlichkeit $> 1-\varepsilon$, daß die Antwort richtig ist. Ist aber der obige Bruch $\leqq 1-\varepsilon$, $\geqq \varepsilon$, so ist keine Antwort mit solcher Sicherheit möglich. Sei die Menge der ν, für die unser Bruch $> 1-\varepsilon$ oder $< \varepsilon$ ist, $Z(n, \varepsilon)$. Dann ist die a priori Wahrscheinlichkeit dafür, daß wir eine Antwort auf Grund des Meßresultates geben können, deren Wahrscheinlichkeit richtig zu sein $> 1-\varepsilon$ ist,

$$\frac{\sum_{\nu \text{ in } Z(n,\varepsilon)} \mu(N_\nu^{(n)})}{\mu(\Omega)} = w(n, \varepsilon).$$

Ist nun ein $\delta > 0$ gegeben, und wird für jedes hinreichend große n $w(n, \varepsilon) > 1-\delta$, so können wir sagen: ist die Messung hinreichend genau, so wird die a priori Wahrscheinlichkeit dafür, daß wir auf Grund der Messung die Frage „gehört P zu $\mathcal{A}$?“ mit einer Wahrscheinlichkeit $> 1-\varepsilon$ beantworten können werden, ihrerseits $> 1-\delta$.

Wenn also zu jedem $\varepsilon > 0$, $\delta > 0$ ein $n_0 = n_0(\varepsilon, \delta)$ existiert, so daß Obiges für alle $n \geqq n_0$ gilt, so können wir sagen: die Frage „gehört P zu $\mathcal{A}$“ kann mit beliebiger Sicherheit durch Messungen entschieden werden. Wir wollen zeigen, daß diese mathematische Forderung in der Tat erfüllt ist.

2. Nehmen wir an, sie wäre es nicht. Dann gäbe es zwei $\varepsilon > 0$, $\delta > 0$, so daß für unendlich viele n $w(n, \varepsilon) \leqq 1-\delta$ ist, etwa für $n = n_1, n_2, \cdots, (n_1 < n_2 < \cdots)$. Die Menge $\Omega - \sum_{\nu \text{ in } Z(n_\varrho, \varepsilon)} N_\nu^{(n_\varrho)}$ heiße M_ϱ, nach

[33] Um wohlbekannte Eigenheiten der wirklichen Messung wiederzugeben, könnten wir diese Entscheidungen auch nur mit gewissen Wahrscheinlichkeiten (die gegen 1 bzw. 0 streben) fällen, u. ä. Dies würde aber das Schlußresultat nicht ändern.

Annahme ist $\mu(M_\varrho) \geqq \delta \cdot \mu(\Omega)$, dabei sind alle $M_\varrho \subseteqq \Omega$, $\mu(\Omega)$ endlich. Nach dem bekannten Satze von Arzelà[34] gilt dann für die Menge $\bar{M}$ aller P, die unendlich vielen M_ϱ angehören, auch $\mu(\bar{M}) \geqq \delta \cdot \mu(\Omega)$. Für ein P von $\bar{M}$ ist für unendlich viele ϱ P in M_ϱ, d. h. P in einem $N_\nu^{(n_\varrho)}$, aber in keinem $N_\nu^{(n_\varrho)}$, ν aus $Z(n_\varrho, \varepsilon)$. D. h. es ist P in $N_\nu^{(n_\varrho)}$, ν nicht in $Z(n_\varrho, \varepsilon)$. Somit ist für unendlich viele n P in $N_\nu^{(n)}$, ν nicht in $Z(n, \varepsilon)$, also

$$\varepsilon \leqq \frac{\mu(\Lambda \cdot N_\nu^{(n)})}{\mu(N_\nu^{(n)})} \leqq 1 - \varepsilon.$$

Wenn wir dasjenige ν, für welches P zu $N_\nu^{(n)}$ gehört, $\nu(n, P)$ nennen, so ist demnach $\lim\limits_{n\to\infty} \frac{\mu(\Lambda \cdot N_{\nu(n,P)}^{(n)})}{\mu(N_{\nu(n,P)}^{(n)})}$ entweder nicht vorhanden, oder $\neq 0, 1$. Dieser Limes kann also nicht überall, bis auf eine Menge vom Maße 0 existieren und $= 0, 1$ sein.

Aber nach einem Satze von Lebesgue ist er überall in Λ bis auf eine Menge vom Maße 0 vorhanden und gleich 1[35], und wenn wir denselben Satz auf $\Omega - \Lambda$ anwenden, so sehen wir, daß er überall in $\Omega - \Lambda$ bis auf eine Menge vom Maße 0 vorhanden und gleich 0 ist. Das steht mit dem vorhin Bewiesenen im Widerspruch.

3. Soweit wir beurteilen können, deckt sich unser Standpunkt — Einführung eines Maßes in Ω, und Beurteilung aller Wahrscheinlichkeiten mit seiner Hilfe — mit dem, was bei physikalischen Messungen wirklich geschieht. Akzeptiert man ihn aber, so ist es unmöglich, der in 2 bewiesenen Möglichkeit, die Zugehörigkeit zu meßbaren Mengen zu entscheiden, auszuweichen. Insbesondere ist es wichtig, daß Mengen vom Maße 0 unbeachtet bleiben müssen, und Eigenschaften, die nur in solchen gelten, als unbeobachtbar anzusehen sind.

Daß die Stetigkeit dabei gar keine Rolle spielte, liegt daran, daß es von diesem Standpunkte aus sozusagen keine unstetige Funktionen gibt: nach einem Satze von Lusin ist ja für jede meßbare Funktion und jedes

---

[34] Vgl. z. B. de la Vallée Poussin, Cours d'Analyse Infinitésimale, Paris 1909, S. 68—69.

[35] Ein Beweis, dem ohne wesentliche Änderungen die hier erforderliche Allgemeinheit zukommt, findet sich bei Carathéodory, „Reelle Funktionen", S. 492—497, Satz 3. Die dort auftretende Funktion $f(P)$ ist $\begin{cases} = 1, \text{ für } P \text{ in } \Lambda \\ = 0, \text{ sonst} \end{cases}$ zu setzen. Der Beweis a. a. O. beruht auf dem „Überdeckungssatz von Vitali", S. 299—307. Für die dortigen $\sigma_n(P)$ sind die $N_{\nu(n,P)}^{(n)}$ einzusetzen. Da in Ω keine Parallelepipede definiert wurden, brauchen die dortigen Prämissen bez. des Verhältnisses der $\sigma_n(P)$ zu Parallelepipeden nicht zu gelten; da aber je zwei $N_{\nu(n,P)}^{(n)}$ (n, P variabel!) fremd sind, oder eines das andere umfaßt, passen sie doch ins Gefüge des Beweises. Sogar etwas besser, als die $\sigma_n(P)$, denn die auf S. 301 ff. erfolgende Einführung der Würfel von dreifacher Kantenlänge wird überflüssig.

$\delta > 0$ eine Menge vom Maße $< \delta$ angebbar, außerhalb welcher die Funktion gleichmäßig stetig ist[36].

II. Ergodische Strömungen.

1. Ω sei ein Raum nach D_1, S_t eine allgemeine Strömung in ihm (D_2'), U_t die Koopmansche Operatorenschar, $U_t = e^{itA}$, A hypermaximal, $E(\lambda)$ seine Zerlegung der Einheit (Einleitung 3).

Wegen der Unitarität von U_t und der Schwarzschen Ungleichheit ist

$$|(U_t f, g)| \leqq \|U_t f\| \cdot \|g\| = \|f\| \cdot \|g\|,$$

also auch

$$\left| \frac{1}{s-t} \int_t^s (U_\tau f, g)\, d\tau \right| \leqq \|f\| \cdot \|g\|.$$

Infolgedessen kann ein Operator $X_{t,s}$ durch

$$(X_{t,s} f, g) = \frac{1}{s-t} \int_t^s (U_\tau f, g)\, d\tau \qquad (t < s)$$

definiert werden[37].

Für jedes meßbare Λ endlichen Maßes können wir $\chi_\Lambda(P) \begin{cases} = 1, & \text{für } P \text{ in } \Lambda \\ = 0, & \text{sonst} \end{cases}$ bilden, und es gehört zum Hilbertschen Raume. Es wird nun einerseits

$$(X_{t,s} \chi_\Lambda, \chi_M) = \int_M X_{t,s} \chi_\Lambda(P)\, dv_P,$$

und andererseits

$$\begin{aligned} (X_{t,s} \chi_\Lambda, \chi_M) &= \frac{1}{s-t} \int_t^s (U_\tau \chi_\Lambda, \chi_M)\, d\tau \\ &= \frac{1}{s-t} \int_t^s \left(\int_M \chi_{S_{-\tau}\Lambda}(P)\, dv_P \right) d\tau, \end{aligned}$$

also nach dem Satz von Fubini (vgl. Anm. [10])

$$= \frac{1}{s-t} \int_M \left(\int_t^s \chi_{S_{-\tau}\Lambda}(P)\, d\tau \right) dv_P = \frac{1}{s-t} \int_M Z(\Lambda, P, t, s)\, dv_P.$$

Da dies für alle M gilt, ist $X_{t,s} \chi_\Lambda(P) = \frac{1}{s-t} Z(\Lambda, P, t, s)$[38].

---

[36] C. R. 154, 1912, S. 1688, oder Sierpinski, Fund. Math. 3, 1922, S. 320. Die dort für das gewöhnliche Lebesguesche Maß geführten Beweise sind für alle l-Maße gültig.

[37] Vgl. a. a. O. Anm. [27], S. 72, oder den analogen Beweis in E, S. 112 oben. Die zu verwendenden Bezeichnungen sind auch nach E.

[38] Eigentlich braucht dies bloß bis auf eine P-Menge vom Maße 0 zu gelten, aber $X_{t,s} \chi_\Lambda$ ist als Element des Hilbertschen Raumes ohnehin nur bis auf eine solche definiert.

Wir berechnen nun $\| X_{t,s} f \|$. Es ist:

$$\begin{aligned}
\| X_{t,s} f \|^2 = \| (X_{t,s} f,\, X_{t,s} f) \| &= \frac{1}{s-t} \int_t^s (U_\tau f,\, X_{t,s} f)\, d\tau \\
&= \frac{1}{(s-t)^2} \int_t^s \int_t^s (U_\tau f,\, U_\sigma f)\, d\tau\, d\sigma \\
&=^{39} \frac{1}{(s-t)^2} \int_t^s \int_t^s (U_{\tau-\sigma} f, f)\, d\tau\, d\sigma \\
&=^{40} \frac{1}{(s-t)^2} \int_{-(s-t)}^{+(s-t)} \int_{2t-|x|}^{2s+|x|} (U_x f, f)\, \frac{dx\, dy}{2} \\
&= \frac{1}{(s-t)^2} \int_{-(s-t)}^{+(s-t)} (s-t-|x|)\, (U_x f, f)\, dx.
\end{aligned}$$

Nun setzen wir nach Einleitung 3 die $E(\lambda)$ für die U_x ein und bemerken, daß die Integrationsreihenfolgen-Vertauschung wegen der absoluten Konvergenz aller auftretenden Integrale zulässig ist. So wird fortlaufend:

$$\begin{aligned}
&= \frac{1}{(s-t)^2} \int_{-(s-t)}^{+(s-t)} \int_{-\infty}^{+\infty} (s-t-|x|)\, e^{i\lambda x}\, dx\, d(E(\lambda) f, f) \\
&= \frac{1}{(s-t)^2} \int_{-\infty}^{+\infty} \left(\int_{-(s-t)}^{+(s-t)} (s-t-|x|)\, e^{i\lambda x} dx \right) d(E(\lambda) f, f) \\
&= \frac{2}{(s-t)^2} \int_{-\infty}^{+\infty} \left(\int_0^{s-t} (s-t-x) \cos(\lambda x)\, dx \right) d(E(\lambda) f, f) \\
&= \frac{2}{(s-t)^2} \int_{-\infty}^{+\infty} \frac{1-\cos(s-t)\lambda}{\lambda^2}\, d(E(\lambda) f, f) \\
&= \int_{-\infty}^{+\infty} \left(\frac{\sin \frac{s-t}{2} \lambda}{\frac{s-t}{2} \lambda} \right)^2 d(E(\lambda) f, f).
\end{aligned}$$

Der Integrand ist $\geqq 0$ und der Ausdruck hinter dem d-Zeichen monoton nichtfallend, also können wir das Integral folgendermaßen nach oben abschätzen: Es ist gleich $\int_{-\infty}^{-\varepsilon} + \int_{+\varepsilon}^{+\infty}$ plus $\int_{-\varepsilon}^{+\varepsilon}$. Vergrößern wir den Integranden hierin zu $\dfrac{1}{\left(\frac{s-t}{2} \varepsilon\right)^2}$ bzw. 1, und ersetzen wir den ersten Integrationsbereich dann durch $\int_{-\infty}^{+\infty}$:

[39] Wegen $U_\sigma^* U_\tau = U_\sigma^{-1} U_\tau = U_{\tau-\sigma}$.

[40] Man substituiere $\sigma - \tau = x$, $\tau + \sigma = y$.

$$\|X_{t,s}f\|^2 \leqq \frac{4}{(s-t)^2\varepsilon^2}\int_{-\infty}^{+\infty} d(E(\lambda)f,f)+\int_{-\varepsilon}^{+\varepsilon} d(E(\lambda)f,f)$$
$$= \frac{4}{(s-t)^2\varepsilon^2}(f,f)+((E(+\varepsilon)-E(-\varepsilon))f,f).$$

Somit ist $\limes_{s-t\to+\infty}\sup \|X_{t,s}f\|^2 \leqq ((E(+\varepsilon)-E(-\varepsilon))f,f)$, also wird, wenn für $\varepsilon\to 0$ $(E(+\varepsilon)-E(-\varepsilon))f\to 0$, $\limes_{s-t\to+\infty}\|X_{t,s}f\|^2=0$, d. h. $X_{t,s}f\to 0$ für $s-t\to+\infty$[41].

Als Projektionsoperator, der monoton nichtfallend von ε abhängt, hat $E(+\varepsilon)-E(-\varepsilon)$ einen Limes für $\varepsilon\to 0$: E_0[42]. Wir sehen also, daß aus $E_0f=0$ $X_{t,s}f\to 0$ für $s-t\to+\infty$ folgt. Ist hingegen $E_0f=f$, so ist wegen $E(+\varepsilon)-E(-\varepsilon)\geqq E_0$ $E(+\varepsilon)f=f$, $E(-\varepsilon)f=0$, d. h. $E(\lambda)f\begin{cases}=f, & \text{für } \lambda>0\\ =0, & \text{für } \lambda<0\end{cases}$. Also für alle g $(U_tf,g)=\int_{-\infty}^{+\infty} e^{it\lambda}(E(\lambda)f,g)$ $=(f,g)$, d. h. $U_tf=f$.

Sei $\mathfrak{M}$ die zu E_0 gehörige abgeschlossene Linearmannigfaltigkeit. In ihr gilt stets $E_0f=f$, also $U_tf=f$, also $X_{t,s}f=f$. Für zu $\mathfrak{M}$ orthogonale f gilt $E_0f=0$, also $X_{t,s}f\to 0$ $(s-t\to+\infty)$. Somit ist für beide Arten von f $X_{t,s}f\to E_0f$ $(s-t\to+\infty)$, also auch für alle $f=f_1+f_2$, f_1 aus $\mathfrak{M}$, f_2 orthogonal zu $\mathfrak{M}$ — d. i. aber jedes f. Wir haben also im Sinne der starken Operatorenkonvergenz $X_{t,s}\to E_0$ für $s-t\to+\infty$[43].

Da aus $E_0f=f$ $U_tf=f$ für alle t folgt, und aus $U_tf=f$ $X_{t,s}f=f$, also $E_0f=f$, können wir $\mathfrak{M}$ auch so charakterisieren: es ist die Menge der gemeinsamen Lösungen aller $U_tf=f$.

2. b), c) aus Einleitung 2 besagen, wie wir nunmehr wissen, dieses: $X_{t,s}\chi_\Lambda\to(\chi_\Lambda,\varphi_0)\cdot\varphi_0$ bzw. $(X_{t,s}\chi_\Lambda,\chi_M)\to(\chi_\Lambda,\varphi_0)\cdot(\varphi_0,\chi_M)$ $(s-t\to+\infty)$, wobei $\varphi_0=\varphi_0(P)=$ konstant ist, und zwar $=\frac{1}{\sqrt{\mu(\Omega)}}$, wenn $\mu(\Omega)$ endlich ist, und $=0$, wenn dieses unendlich ist (also $\|\varphi_0\|=1$ bzw. $\varphi_0=0$). Wir können also auch sagen: $X_{t,s}f\to P_{[\varphi_0]}f$ bzw. $(X_{t,s}f,g)\to(P_{[\varphi_0]}f,g)$, falls f,g die Form χ_Λ haben. Dann muß aber diese Gleichung auch für alle Linearaggregate der χ_Λ gelten und für deren Häufungspunkte[44], und

[41] Unter $\to 0$ verstehen wir im Hilbertschen Raume die starke Konvergenz, d. h. $\|\cdots\|\to 0$. Vgl. A S. 378.

[42] Vgl. hierfür, wie fürs Folgende, die in E, S. 74—78 entwickelte Theorie der Projektionsoperatoren. Gegenwärtig ist Satz 19 dortselbst gemeint sowie S. 91.

[43] Vgl. A, S. 381—384.

[44] Wegen $|(X_{t,s}f,g)|\leqq\|f\|\cdot\|g\|$ sind die $X_{t,s}$ gleichmäßig beschränkt. Vgl. etwa E, S. 73, Satz 12.

dies sind sämtliche f[45]. Somit besagen b) bzw. c), daß die obigen Gleichungen für alle f, g gelten, d. h., daß $X_{t,s}$ im Sinne der starken bzw. der schwachen Operatorenkonvergenz (vgl. Anm.[43]) gegen $P_{[\varphi_0]}$ strebt. Dies gilt übrigens auch dann, wenn b) bzw. c) nicht für alle meßbaren $\mathfrak{A}$, M endlichen Maßes formuliert werden, sondern nur für eine solche Menge von $\mathfrak{A}$, M, durch die alle obigen approximiert werden können — etwa alle Summen endlich vieler Kugeln oder Parallelepipede (vgl. a. a. O. Anm.[45]). Ja, wegen der Additivität dieser Eigenschaften genügen die Parallelepipede selbst.

Andererseits wissen wir aber, daß $X_{t,s} \to E_0$ (stark). Somit besagen b), c) dasselbe: $E_0 = P_{[\varphi_0]}$, oder, da $\mathfrak{M}$ zu E_0 gehört: $\mathfrak{M} = [\varphi_0]$. Oder: die gemeinsamen Lösungen der $U_t f = f$ sind die $c\varphi_0$, d. h. die Konstanten und nur diese[46].

Daß $f(P)$ Lösung aller $U_t f = f$ ist, bedeutet: für jedes t gilt $f(S_t P) = f(P)$ bis auf eine P-Menge vom Maße 0. Diese f sind die „Integrale" der Strömung. Integrale mit endlichem Absolutwertquadrat-Integral, d. h. im Hilbertschen Raume, nennen wir „H-Integrale". Wir haben also gezeigt:

SATZ 1. b), c) *aus Einleitung* 2 *sind einander gleichwertig, ferner ist es dasselbe, ob man sie für alle meßbaren* $\mathfrak{A}$, M *endlichen Maßes formuliert, oder nur für Parallelepipede, oder Summen endlich vieler Kugeln, o. ä. Für ihre Gültigkeit, d. h. für die Ergodizität der Strömung, ist notwendig und hinreichend, daß deren einzige H-Integrale die Konstanten seien.*

Damit ist der „Ergodensatz" der klassischen Mechanik bewiesen (vgl. dazu Anm.[3], [19]).

Es ist bei ergodischen Ω noch zweckmäßig zu unterscheiden, ob $\mu(\Omega)$ endlich oder unendlich ist, d. h. ob der Limes von $Z(\mathfrak{A}, P, t, s)$ nicht identisch Null ist, oder es ist. Wir nennen Ω dementsprechend eigentlich-ergodisch bzw. Null-ergodisch.

3. Für beliebige, also u. U. nicht ergodische Strömungen können wir die Schlußformel von 1 wieder dazu benutzen, um $\operatorname{limes}_{s-t\to+\infty} \frac{1}{s-t} Z(\mathfrak{A}, P, t, s)$ direkt zu berechnen, dies geschah a. a. O. Anm.[27], S. 75–77. Wir schlagen hier lieber einen indirekten Weg ein, der die in Einleitung 4 angekündigte Zerlegung in ergodische Strömungen mit ergibt.

Ein meßbares $\mathfrak{A}$ endlichen Maßes, welches mit jedem $S_t \mathfrak{A}$ bis auf eine Menge vom Maße 0 (die natürlich von t abhängt) zusammenfällt, heiße eine $\mathfrak{A}$-Menge. Sie ist auch dadurch charakterisierbar, daß $\chi_{\mathfrak{A}}$ $\left(\chi_{\mathfrak{A}}(P) \begin{cases} = 1 \text{ für } P \text{ in } \mathfrak{A} \\ = 0 \quad \text{sonst} \end{cases}\right)$ ein H-Integral ist, d. h. zu $\mathfrak{M}$ gehört. Sei

[45] Vgl. E, S. 109–111.

[46] Damit die Konstante a zum Hilbertschen Raume gehöre, muß $\int_\Omega |a|^2 dv_P = |a|^2 \mu(\Omega)$ endlich sein, bei unendlichem $\mu(\Omega)$ also $a = 0$.

umgekehrt f ein beliebiges H-Integral. Es ist starker Häufungspunkt solcher $g(P) = G(f(P))$, deren jede nur endlich vieler verschiedener Werte fähig ist, und diese ihrerseits Linearaggregate endlich vieler solcher $h(P) = H(f(P))$, die nur die Werte 0, 1 annehmen (vgl. a. a. O. Anm. [45]). Diese sind also $\chi_\Lambda(P)$-Funktionen und ihrer Form nach (H)-Integrale, also sind ihre Λ Λ-Mengen. Daher spannen die χ_Λ aller Λ-Mengen die abgeschlossene Linearmannigfaltigkeit $\mathfrak{M}$ auf. Dasselbe gilt, wenn wir uns auf ein Teilsystem der Λ-Mengen beschränken, das in diesen überall dicht liegt, und es gibt eine Folge $\Lambda_1, \Lambda_2, \cdots$, die dies leistet (vgl. a. a. O. Anm. [45]). Schließlich können wir diese Λ_n, durch Änderungen um Mengen vom Maße 0, alle Borelsch machen.

Die Mengenoperationen $\Lambda + M$, $\Lambda \cdot M$, $\Lambda - M$, $\lim_{n\to\infty} M_n = M_1 \cdot M_2 \cdot \cdots$ (mit $M_1 \supset M_2 \supset \cdots$) führen aus dem Kreise der Borelschen Λ-Mengen nicht hinaus. Infolgedessen kann ein Beweis, den der Verfasser in einem anderen Zusammenhange verwendete, wörtlich übertragen werden[47], und er ergibt:

Sei A das Intervall $0 \leqq x < 1$, $A' = \{x^{(1)}, x^{(2)}, \cdots\}$ eine in ihm überall dichte abzählbare Menge. Jedem x von A' ist eine Borelsche Λ-Menge $\Lambda(x)$ zugeordnet, derart daß

α) Jedes Λ_n entsteht durch die Mengenoperationen $\Lambda + M$, $\Lambda \cdot M$, $\Lambda - M$ aus einigen $\Lambda(x)$.

β) Aus $x \leqq y$ (in A') folgt $\Lambda(x) \subseteq \Lambda(y)$[48].

γ) Aus $x_1 > x_2 > \cdots > x$, $x_n \to x$ (für $n \to \infty$, alles in A') folgt $\lim_{n\to\infty} \Lambda(x_n) = \Lambda(x_1) \cdot \Lambda(x_2) \cdot \cdots = \Lambda(x)$[48].

Sei ferner $K_1, K_2, \cdots$ ein „gleichwertiges Umgebungssystem“ (vgl. a. a. O. Anm. [45]), wir konstruieren ebenso diese Schar:

Jedem ξ von A' ist eine Borelsche Menge $K(\xi)$ zugeordnet, die α), β), γ) ebenfalls erfüllt, jedoch derart, daß durch α) die $\Omega, K_1, K_2, \cdots$ erzeugt werden.

Wir wollen die Mengensummen und -durchschnitte mit Σ bzw. Π bezeichnen. Es sei

$$\sum_{x \text{ in } A'} \Lambda(x) = \Omega_1, \qquad \Omega - \Omega_1 = \Omega_2.$$

Da jedes $\Lambda(x) \subseteq \Omega_1$ ist, gilt dasselbe für jedes Λ_n, also für jede Λ-Menge bis auf eine Menge vom Maße 0. Jedes $\Lambda(x)$ ist gegenüber jedem S_t in-

[47] A, S. 401—403, an Stelle der dortigen Projektionsoperatoren sind Λ-Mengen zu setzen, für die $+, \cdot, -$ ja ebenfalls definiert sind. Insbesondere sind $E_1, E_2, \cdots$ durch $\Lambda_1, \Lambda_2, \cdots$, die $G(\lambda)$ durch die $\Lambda(x)$, $x = \lambda + 1$, zu ersetzen. A' ist irgendeine in $0 \leqq x < 1$ überall dichte Folge, die die dortigen ϱ enthält, etwa die der rationalen Zahlen.

[48] Man beachte, daß die Konstruktionen von a. O. Anm. [47] die Mengenbeziehungen $\subseteq$ und $=$ genau ergeben, und nicht bloß bis auf Mengen vom Maße 0.

variant bis auf eine Menge vom Maße 0, also sind es auch Ω_1, Ω_2. Hätte Ω_2 ein endliches Maß, so wäre es $\mathcal{A}$-Menge, also hätte es, da es zu Ω_1 fremd ist, das Maß 0. D. h.: $\mu(\Omega_2) = 0$ oder ∞. Im ersten Fall brauchen wir Ω_2 nicht zu beachten. Im zweiten ist S_t auch in ihm eine allgemeine Strömung, und jede seiner $\mathcal{A}$-Mengen hat (da sie fremd zu Ω_1 ist) das Maß 0. Ist f ein H-Integral in Ω_2, so ist die Menge $|f(P)| > \varepsilon$ $(\varepsilon > 0)$ eine $\mathcal{A}$-Menge, also vom Maße 0, $\varepsilon = 1, \frac{1}{2}, \frac{1}{3}, \cdots$ ergibt: $f(P) = 0$ bis auf eine Menge vom Maße 0. Somit ist Ω_2 Null-ergodisch[49].

Wir definieren nun für jedes P von Ω_1 einen Punkt $[x, \xi]$ im Quadrat $A \times A = \bar{A}$, d. i. $0 \leqq x, \xi < 1$: $x = \underset{y \text{ in } A'}{\text{finis inf}} (P \text{ in } \mathcal{A}(y))$, $\xi = \underset{\eta \text{ in } A'}{\text{finis inf}}$ $(P \text{ in } K(\eta))$.

Wenn P, P' $[x, \xi]$ bzw. $[x', \xi']$ entsprechen, so folgt aus $\xi = \xi'$ $P = P'$. Denn durch $\mathcal{A} + M$, $\mathcal{A} \cdot M$, $\mathcal{A} - M$ entstehen aus den $K(\eta)$ nur Mengen $(K(\eta_2) - K(\eta_1)) + \cdots + (K(\eta_{2\nu}) - K(\eta_{2\nu-1}))$ $(\eta_1 < \eta_2 < \cdots < \eta_{2\nu-1} < \eta_{2\nu}$, nach $\beta)$, damit P zu einer solchen gehöre, ist die Zugehörigkeit von ξ zu einem $K(\eta_{2\mu}) - K(\eta_{2\mu-1})$ $(\mu = 1, \cdots, \nu)$ charakteristisch, d. h. $\eta_{2\nu-1} < \xi \leqq \eta_{2\nu}$. Wegen $\xi = \xi'$ gehören also P, P' zu denselben unter diesen Mengen, also zu denselben K_n, für $P \neq P'$ müßte es aber ein K_n geben, das P enthält und P' nicht. Unsere Abbildung $P \to [x, \xi]$ ist also eineindeutig, selbst ihre Projektion auf der ξ-Achse ist es noch. Diese Abbildung möge Ω_1 auf B abbilden.

Im Produktraume $\Omega_1 \times \bar{A}$ ist die Menge der $[P, x, \xi]$, für die $[x, \xi]$ gerade das Bild von P ist, gleich

$$\prod_{k=2}^{\infty} \sum_{m,n=1}^{k-1} \bar{M}(x_m^{(k)}, x_n^{(k)}; x_{m+1}^{(k)}, x_{n+1}^{(k)}),$$

falls $x_1^{(k)}, \cdots, x_k^{(k)}$ die der Größe nach geordneten Zahlen $x^{(1)}, \cdots, x^{(k)}$ sind (die k ersten Elemente von A'), und $\bar{M}(y, \eta; z, \zeta)$ die Menge derjeniger $[P, x, \xi]$ ist, für welche entweder $y < x \leqq z$, $\eta < \xi \leqq \zeta$ gilt und P in $(\mathcal{A}(z) - \mathcal{A}(y)) \cdot (K(\zeta) - K(\eta))$, oder aber $y < x \leqq z$, $\eta < \xi \leqq \zeta$ nicht gilt. Da die $\bar{M}(y, \eta; z, \zeta)$ offenbar Borelsch sind, ist es auch die obengenannte Menge. Ist $M \subseteq \Omega_1$ Borelsch, so ist es auch die Menge der $[P, x, \xi]$ mit P in M, und auch der Durchschnitt dieser Menge mit jener: d. h. die Menge der $[P, x, \xi]$, P in M, $[x, \xi]$ das Bild von P. Ihre Projektion auf den $[x, \xi]$-Raum $A \times A = \bar{A}$ ist nach dem w. o. Gesagten eindeutig, also Borelsch[50], dies ist aber das Bild von M. Wir können insbesondere $M = \Omega_1$ setzen, und erhalten: das durch $P \to [x, \xi]$ vermittelte Bild $\bar{B}$ von Ω_1, sowie dasjenige jeder Borelschen Teilmenge von Ω_1, ist eine Borelsche Menge.

[49] Die Zerlegung von Ω in Ω_1, Ω_2 verdankt der Verfasser einer Mitteilung von Herrn M. H. Stone.

[50] Vgl. z. B. Hausdorff, „Mengenlehre", S. 212, Satz 4; oder Lusin, Fund. Math. 10, 1928, S. 1.

Sei $\overline{A}(y, \eta; z, \zeta)$ das Quadrat $y < x \leqq z$, $\eta < \xi \leqq \zeta$, sein Urbild (d. h. dasjenige seines in $\overline{B}$ liegenden Teiles) ist das meßbare

$$(\Lambda(z) - \Lambda(y)) \cdot (K(\zeta) - K(\eta)),$$

dessen Maß

$$\mu\,[(\Lambda(z) - \Lambda(y)) \cdot (K(\zeta) - K(\eta))] = \overline{\mu}(y, \eta; z, \zeta).$$

Dies hat die Konsequenz, daß aus

$$\overline{A}(y, \eta; z, \zeta) \cdot \overline{B} \subseteq \sum_1^\infty{}^n \overline{A}(y^{(n)}, \eta^{(n)}; z^{(n)}, \zeta^{(n)}) \cdot \overline{B}$$

$$\overline{\mu}(y, \eta; z, \zeta) \leqq \sum_1^\infty{}^n \overline{\mu}(y^{(n)}, \eta^{(n)}; z^{(n)}, \zeta^{(n)})$$

folgt. Also hat

$$\overline{A}(y, \eta; z, \zeta) \cdot \overline{B} \subseteq \overline{A}(y', \eta'; z', \zeta') \cdot \overline{B}$$

die Folge

$$\overline{\mu}(y, \eta; z, \zeta) \leqq \overline{\mu}(y', \eta'; z', \zeta'),$$

und aus $=$ statt $\subseteq$ folgt durch Vertauschung beider Seiten $=$ statt $\leqq$. Ferner zieht

$$\overline{A}(y, \eta; z, \zeta) \subseteq \sum_1^\infty{}^n \overline{A}(y^{(n)}, \eta^{(n)}; z^{(n)}, \zeta^{(n)})$$

erst recht

$$\overline{\mu}(y, \eta; z, \zeta) \leqq \sum_1^\infty{}^n \overline{\mu}(y^{(n)}, \eta^{(n)}; z^{(n)}, \zeta^{(n)})$$

nach sich. Schließlich ist $\overline{\mu}(y, \eta; z, \zeta)$ in allen vier Variablen nach rechts halbstetig, weil es $\Lambda(x)$, $K(\xi)$ auch sind.

Wir können nunmehr für die Teilmengen $\overline{C}$ von $\overline{A}$ definieren:

$\overline{\mu}^*(\overline{C})$ *sei die untere Grenze aller*

$$\sum_1^\infty{}^n \overline{\mu}(y^{(n)}, \eta^{(n)}; z^{(n)}, \zeta^{(n)})$$

mit

$$\overline{C} \subseteq \sum_1^\infty{}^n \overline{A}(y^{(n)}, \eta^{(n)}; z^{(n)}, \zeta^{(n)}).$$

Aus dem soeben Gezeigten folgt, daß dies ein l-Maß im Quadrat $\overline{A}$ (in bezug auf die gewöhnliche Topologie von $\overline{A}$) ist. Ferner werden wir sofort beweisen: wenn $\overline{C}$ $\overline{\mu}$-meßbar ist, so haben $\overline{C}$ und $\overline{C} \cdot \overline{B}$ dasselbe Maß. Da $\overline{B}$ Borelsch ist, ist es meßbar, $\overline{A} - \overline{B}$ ist es auch, und es ist

$$\bar{\mu}(\bar{A} - \bar{B}) = \mu((\bar{A} - \bar{B}) \cdot \bar{B}) = 0.$$

$[x, \xi] \to P$ bildet $\bar{B}$ eineindeutig auf Ω_1 ab. Wenn $\bar{C} = \bar{A}(y, \eta; z, \zeta)$ ist, so ist das Urbild von $\bar{C} \cdot \bar{B}$ meßbar, und sein μ-Maß ist $= \bar{\mu}(\bar{C} \cdot \bar{B}) = \bar{\mu}(\bar{C})$ — in der Tat ist das Urbild gleich $(A(z) - A(y)) \cdot (K(\zeta) - K(\eta))$, sein Maß $\mu(y, \eta; z, \zeta)$, und dies ist, wie aus der auf S. 604 oben bewiesenen Formel folgt, $= \bar{\mu}(\bar{C} \cdot \bar{B}) = \bar{\mu}(\bar{C})$. Dasselbe gilt also, wenn $\bar{C}$ die Summe endlich vieler fremder $A(y, \eta; z, \zeta)$ ist, oder Limes einer aufsteigenden Folge solcher, d. h. für jedes offene $\bar{C}$. Von diesen überträgt es sich wiederum durch Limesbildung auf- oder absteigender Folgen auf alle Borelschen $\bar{C}$. Hat $\bar{C}$ das $\bar{\mu}$-Maß 0, so hat es auch $\bar{C} \cdot \bar{B}$, und es ist Teil eines ebensolchen Borelschen $\bar{C}'$; das Urbild von $\bar{C}' \cdot \bar{B}$, also auch dasjenige von $\bar{C} \cdot \bar{B}$, hat dann auch das μ-Maß 0. Ist schließlich $\bar{C}$ bloß $\bar{\mu}$-meßbar, so ist es Summe einer Borelschen Menge und einer vom $\bar{\mu}$-Maß 0; da die obige Behauptung für beide gilt, gilt sie auch für $\bar{C}$.

Für die $\bar{C} \subseteqq \bar{B}$ haben wir insbesondere gewonnen: Ist $\bar{C}$ $\bar{\mu}$-meßbar, so ist sein $[x, \xi] \to P$-Bild μ-meßbar und vom selben Maß. Aus Satz 2 der vorhergehenden Abhandlung des Verfassers folgt alsdann, daß $[x, \xi] \to P$ sowie $P \to [x, \xi]$ überhaupt maßtreu sind.

4. Das $\bar{\mu}$-Maß in $\bar{A}$ (oder $\bar{B}$) läßt eine Integraldarstellung zu, die wir angeben wollen.

Wir definieren für alle x, ξ in $0 \leqq x \leqq 1$, $0 \leqq \xi \leqq 1$ (nicht nur für die in $\bar{A}$!)

$$\bar{\mu}(x, \xi) = \bar{\mu}(\bar{A}(0, 0; x, \xi)).$$

Wenn x, ξ zu A' gehören, ist also

$$\bar{\mu}(x, \xi) = \bar{\mu}(0, 0; x, \xi) = \mu(A(x) \cdot K(\xi)).$$

Offenbar sind $\bar{\mu}(x, \xi)$ und $\bar{\mu}(x, \xi'') - \bar{\mu}(x, \xi')$ $(\xi' \leqq \xi'')$ monoton nichtfallend und nach rechts halbstetig in x. Wenn wir also ξ festhalten, x von 0 bis 1 variieren lassen, aber $u = \bar{\mu}(x, 1)$ als unabhängige Variable ansehen, so liegen die Differenzenquotienten von $\bar{\mu}(x, \xi)$ stets im Intervalle 0, 1. Insbesondere ist in den „Konstanzintervallen" von $u = \bar{\mu}(x, 1)$ $\bar{\mu}(x, \xi)$ ebenfalls konstant. Die von $\bar{\mu}(0, 1) = 0$ bis $\bar{\mu}(1, 1) = \mu(\Omega_1)$ variierende Variable u überspringt allerdings die den Unstetigkeitsstellen x von $\bar{\mu}(x, 1)$ entsprechenden „Sprungintervalle" $\bar{\mu}(x-0, 1) \leqq u < \bar{\mu}(x+0, 1) = \bar{\mu}(x, 1)$, aber in jedem dieser Intervalle können wir $\bar{\mu}(x, \xi)$ linear und an den Enden stetig in u neu definieren. Nunmehr ist der Satz von Lebesgue anwendbar[51]: die Derivierte

[51] Vgl. z. B. Carathéodory, „Reelle Funktionen", S. 551 und S. 545.

$$\varrho(x, \xi) = \operatorname*{limes}_{\substack{\varepsilon>0, \delta>0 \\ \varepsilon, \delta \to 0}} \frac{\overline{\mu}(x+\varepsilon, \xi) - \overline{\mu}(x-\delta, \xi)}{\overline{\mu}(x+\varepsilon, 1) - \overline{\mu}(x-\delta, 1)}$$

existiert überall, abgesehen von einer solchen x-Menge, deren u-Menge das gewöhnliche Lebesguesche Maß 0 hat. Die obengenannten „Sprungintervalle" von u brauchen wir nicht zu beachten, denn in den entsprechenden Unstetigkeitsstellen x existiert der obige Limes, da sein Nenner nicht gegen 0 strebt, und im u-Intervalle hat wegen der Linearität die Derivierte denselben Wert. Wenn wir also für Teilmengen X der x-Menge $0 \leqq x \leqq 1$ ein Maß

$$\tilde{\mu}^{(0)*}(X) = \int_X d\,\overline{\mu}(x, 1)$$

einführen[52], das offenbar ein l-Maß ist, so können wir sagen: $\varrho(x, \xi)$ existiert für alle x, außer für eine Menge vom $\tilde{\mu}^{(0)}$-Maße 0.

Weiter ist $\overline{w}(x, \xi)$ das Integral seiner Derivierten:

$$\overline{w}(x, \xi) = \int_0^x \varrho(y, \xi)\, d_y\, \overline{w}(y, 1).$$

Wir lassen nun ξ nur in A' variieren. Alle diese $\varrho(x, \xi)$ existieren gleichzeitig, bis auf eine x-Menge Y_0 vom Maße 0, deren Komplementärmenge in $0 \leqq x \leqq 1$ X_0 heiße. Da für $\xi' \leqq \xi''$ $\varrho(x, \xi'') - \varrho(x, \xi')$ die Derivierte von $\overline{w}(x, \xi'') - \overline{w}(x, \xi')$ ist, also $\geqq 0$, ist $\varrho(x, \xi)$ in ξ monoton nichtfallend. Wir definieren nun für alle ξ

$$\varrho_1(x, \xi) = \operatorname*{limes}_{\substack{\eta \text{ in } A', \eta > \xi \\ \eta \to \xi}} \varrho(x, \eta),$$

$\varrho_1(x, \xi)$ ist in ξ monoton nichtfallend und nach rechts halbstetig. Da $\overline{w}(x, \xi)$ in ξ nach rechts halbstetig ist, wird aus unserer obigen Integralformel durch denselben Grenzprozeß

$$\overline{w}(x, \xi) = \int_0^x \varrho_1(y, \xi)\, d_y\, \overline{w}(y, 1).$$

Wir definieren nun jeden Parameterwert x aus X_0 und für jede Teilmenge Ξ der Menge $0 \leqq \xi \leqq 1$ ein Maß $\tilde{\mu}_x(\Xi) = \int_\Xi d_\xi\, \varrho_1(x, \xi)$ (vgl. das am Ende von Anm. [52] Gesagte), das offenbar ein l-Maß ist. Dabei ist die gewöhnliche Topologie von $0 \leqq \xi \leqq 1$ zugrunde zu legen. Wenn $\overline{C}$ eine Menge $\subseteq \overline{A}$ ist, so sei $(\overline{C})_{(x)}$ die Menge aller ξ mit $[x, \xi]$ in $\overline{C}$. Sei $\overline{C}$ die Menge $\overline{A}$ $(0, 0; x, \xi)$, dann ist:

[52] D. i. das äußere Maß der entsprechenden u-Menge, wenn diese unmeßbar ist, stellt $\int_X$ das „obere Lebesguesche Integral" dar.

$$\overline{\mu}(\overline{C}) = \overline{w}(x, \xi) = \int_0^x \varrho_1(y, \xi)\, d_y\, \overline{w}(y, 1)$$
$$= \int_0^x \left[\int_0^\xi d_\eta\, \varrho_1(y, \eta)\right] d_y\, \overline{w}(y, 1) = \int_0^1 \left[\int_{(\overline{C})_{(y)}} d_\eta\, \varrho_1(y, \eta)\right] d_y\, \overline{w}(y, 1)$$
$$= \int_0^1 \tilde{\mu}_{(y)}\left((\overline{C})_{(y)}\right) d_y\, \overline{w}(y, 1).$$

Wir fragen allgemein: für welche $\overline{C} \subseteq \overline{A}$ gilt die Formel

$$\overline{\mu}(\overline{C}) = \int_0^1 \tilde{\mu}_{(y)}\left((\overline{C})_{(y)}\right) d_y\, \overline{w}(y, 1),$$

wobei auch verlangt wird, daß der Integrand überall, bis auf eine y-Menge vom $\tilde{\mu}^{(0)}$-Maße 0, Sinn hat (nicht bloß als $\tilde{\mu}^*_{(y)}$, sondern als $\tilde{\mu}_{(y)}$, d. h. $(\overline{C})_{(y)}$ soll $\tilde{\mu}_{(y)}$-meßbar sein), und eine $\tilde{\mu}^{(0)}$-meßbare Funktion ist (d. h. das Integral Sinn hat)? Dies ist für die $\overline{C} = \overline{A}(0, 0; x, \xi)$ der Fall, also auch für ihre Differenzen, die $\overline{C} = \overline{A}(x, \xi; y, \eta)$; sodann für die Summen endlich vieler solcher und die Limites aufsteigender Folgen solcher Mengen — d. i. für alle offenen $\overline{C}$. Genau wie in der Betrachtung am Ende von 3. überträgt sich unsere Formel von diesen $\overline{C}$ successiv auf die Borelschen $\overline{C}$, auf die $\overline{C}$ vom $\overline{\mu}$-Maße 0, und schließlich auf alle $\overline{\mu}$-meßbaren $\overline{C}$.

Zum Schluß stellen wir noch dieses fest. Sei X eine Menge in A, $\overline{X}$ die Menge aller $[x, \xi]$, x in X, $0 \leqq \xi < 1$ (in $\overline{A}$). Wenn $\overline{X}$ $\overline{\mu}$-meßbar ist, ergibt unsere Formel sofort: X ist $\tilde{\mu}^{(0)}$-meßbar, $\tilde{\mu}^{(0)}(X) = \overline{\mu}(\overline{X})$. Ist umgekehrt X $\tilde{\mu}^{(0)}$-meßbar, so gibt es zwei Borelsche X', X'' mit $X' \subseteq X \subseteq X''$, $\tilde{\mu}^{(0)}(X'' - X') = 0$, also sind $\overline{X}'$, $\overline{X}''$ auch Borelsch, also meßbar, und es ist $\overline{\mu}(\overline{X}'' - \overline{X}') = 0$. Wegen $\overline{X}' \subseteq \overline{X} \subseteq \overline{X}''$ ist auch $\overline{X}$ meßbar und hat dasselbe $\overline{\mu}$-Maß wie $\overline{X}$, $\overline{X}''$, d. h. wie X', X'', d. h. wie X. Also: X, $\overline{X}$ sind immer gleichzeitig meßbar und haben dasselbe ($\tilde{\mu}^{(0)}$- bzw. $\overline{\mu}$-)Maß. Da $\overline{A} - \overline{B}$ das Maß 0 hat, gilt dasselbe, wenn wir $\overline{X}$ auf $\overline{B}$ einschränken.

5. Wir übertragen nun $\overline{B}$ durch die maßtreue Abbildung $[x, \xi] \to P$ zurück nach Ω_1. Aus der Menge der $[y, \eta]$ mit $y \leqq x$, $0 \leqq \eta \leqq 1$ wird $\Lambda(x)$, aus der mit $y = x$, $0 \leqq \eta \leqq 1$ also

$$M(x) = \prod_{\substack{y \text{ in } A' \\ y > x}} \Lambda(y) - \sum_{\substack{y \text{ in } A' \\ y < x}} \Lambda(y).$$

Wir übertragen dabei auch die in $\overline{A}$ und seinen Teilmengen definierten l-Maße: aus $\overline{\mu}$ in $\overline{B}$ wird, wie wir wissen, μ in Ω_1; $\tilde{\mu}_{(x)}$ übertragen wir auf $M(x)$, wo es $\mu_{(x)}$ heißen soll — all dies sind l-Maße. Immerhin ist bei $\mu_{(x)}$ dieses zu beachten: Bei ihm ist jene Topologie zugrunde zu

legen, die bei $\xi \to$ ins zu $[x, \xi]$ gehörige P aus der gewöhnlichen Topologie von $0 \leqq \xi \leqq 1$ entsteht. Eine andere denkbare Topologie für $M(x)$ ist die, bei der die in $M(x)$ relativ-offenen Mengen (im Sinne der Ω-Topologie, also die $O \cdot M(x)$, O offen in Ω) als offen gelten. Nun sind die Borelschen Mengen der ersteren Topologie aus den Bildern der ξ-Mengen $\xi \leqq \eta$, η in A' erzeugbar, d. h. aus den Mengen $M(x) \cdot K(\eta)$, oder, was nach Konstruktion der $K(\eta)$ dasselbe ist, aus den $M(x) \cdot K_n$. Die Borelschen Mengen der letzteren Topologie entstehen aus den $M(x) \cdot O$, also auch aus den $M(x) \cdot K_n$. Beide Topologien haben also dieselben Borelschen Mengen, da aber für den l-Maßcharakter nur die Borelschen Mengen maßgebend sind (vgl. Anm. [9] in der vorhergehenden Abhandlung des Verfassers), ist $\mu_{(x)}$ auch im Sinne der letzteren Topologie, d. i. der auf $M(x)$ übertragenen Topologie von Ω, ein l-Maß.

Ist $M[X] = \sum_{x \text{ in } X} M(x)$, so ist $M[X]$ das Bild von $\overline{X}$, also sind auch X und $M[X]$ gleichzeitig meßbar, und haben dieselben ($\tilde{\mu}^{(0)}$- bzw. μ-)Maße. Aus unserer $\overline{\mu}(\overline{C})$-Formel aber wird: für jedes meßbare $N \subseteqq \Omega_1$ ist

$$\mu(N) = \int_0^1 \mu_{(x)}(M(x) \cdot N)\, dw(x).$$

Dabei ist $w(x) = w(x, 1) = \tilde{\mu}^{(0)}$ (Menge aller y mit $0 < y \leqq x$). Der Integrand kann für einige x sinnlos sein ($\mu^*_{(x)}$ undefiniert oder $M(x) \cdot N$ unmeßbar), aber die Menge dieser x hat das $\tilde{\mu}^{(0)}$-Maß 0; insbesondere ist $\mu^*_{(x)}$ nur für die x von X_0 definiert, aber diese Komplementärmenge von X_0 in $0 \leqq x \leqq 1$, Y_0, hat das $\tilde{\mu}^{(0)}$-Maß 0.

Untersuchen wir nunmehr das Verhältnis der S_t zu den $M(x)$. Sei T der m-Raum der t, $\tilde{\mu}$ das gewöhnliche Lebesguesche Maß in ihm, $\Omega \times T$ der Produktraum, $\overline{\mu}$ das dort nach D_1 definierte l-Maß (die Verwechslung mit dem $\overline{\mu}$ in 4. droht nicht, da dieses nicht mehr vorkommen wird).

Sei $M \subseteqq \Omega$. Wenn es offen ist, ist sein $[P, t] \to S_t P$-Urbild in $\Omega \times T$ meßbar (d. h. die Funktion $S_t P$ ist meßbar, vgl. Anm. [11]), und von den offenen M überträgt sich dies mühelos auf die Borelschen. Hat M außerdem das Maß 0, so ist für jedes feste t die Menge der P mit $S_t P$ in M vom Maße 0 (S_t ist maßtreu), also ergibt der Satz von Fubini (a. a. O. Anm. [10]), den wir auf das Urbild wegen seiner Meßbarkeit anwenden dürfen: das Urbild selbst hat das Maß 0. Hat M bloß das Maß 0, so ist $M \subseteqq M'$, M' Borelsch und vom Maß 0, also Urbild von $M \subseteqq$ Urbild von M', das vom Maße 0 ist, also das Urbild von M auch. Das Urbild von M ist also meßbar, wenn M Borelsch ist, oder vom Maße 0, also auch für die Summen zweier solcher M, d. h. für jedes meßbare M.

Gehöre x zu A'. $A(x)$ ist meßbar, also auch die Menge der $[P, t]$, P in $A(x)$, ferner das $S_t P$-Urbild: die Menge der $[P, t]$, $S_t P$ in $A(x)$.

Also ist auch die Summe minus der Durchschnitt dieser zwei Mengen meßbar, d. i. die Menge der $[P, t]$, für die P zu $\mathcal{A}(x)$ gehört, aber $S_t P$ nicht, oder umgekehrt. Und auch die Summe dieser Mengen für alle x von A' ist meßbar, sie heiße $\overline{Y}_1$. Nun sei t fest. Die Menge der P, für die $[P, t]$ zu $\overline{Y}_1$ gehört, ist

$$\sum_{x \text{ in } A'} [(\mathcal{A}(x) + S_t^{-1}\mathcal{A}(x)) - (\mathcal{A}(x) \cdot S_t^{-1}\mathcal{A}(x))],$$

also, wie jeder ihrer Addenden, vom Maße 0. Der Satz von Fubini ist anwendbar: $\overline{Y}_1$ hat das Maß 0. $[P, t]$ nicht in $\overline{Y}_1$ bedeutet: P gehört zu denselben $\mathcal{A}(x)$ (x in A'), wie $S_t P$, also auch zu denselben $\boldsymbol{M}(x)$ ($0 \leqq x < 1$), oder: P gehört dann und nur dann zu $\boldsymbol{M}(x)$, wenn $S_t P$ es tut, u. zw. für alle $0 \leqq x < 1$.

Nun definieren wir $S_t P$ um, indem wir es in $\overline{Y}_1$ für undefiniert erklären, sonst aber nicht ändern. Da $\overline{Y}_1$ das Maß 0 hat, bleibt $S_t P$ meßbar; da bei festem t die Menge der P mit $[P, t]$ in $\overline{Y}_1$ das Maß 0 hat, sind die neuen S_t den alten äquivalent. Dabei sind alle $\boldsymbol{M}(x)$ den neuen S_t gegenüber genau invariant. Auf Grund der Ersetzungsmöglichkeit durch eine äquivalente allgemeine Strömung dürfen wir also annehmen, daß die $\boldsymbol{M}(x)$ von vornherein genau invariant waren.

Da jedes $\boldsymbol{M}(x)$ genau S_t-invariant ist, ist es auch jedes $\boldsymbol{M}[X]$. Ist X meßbar, $\tilde{\mu}^{(0)}(X)$ endlich, so ist auch $\boldsymbol{M}[X]$ meßbar, $\mu(\boldsymbol{M}[X])$ endlich, d. h. $\boldsymbol{M}[X]$ $\mathcal{A}$-Menge. Umgekehrt ist jedes $\mathcal{A}(\eta)$ ein $\boldsymbol{M}[X]$ (X ist die Menge $0 \leqq \xi \leqq \eta$) mit meßbarem X; also auch alles, was durch $\mathcal{A} + \boldsymbol{M}$, $\mathcal{A} \cdot \boldsymbol{M}$, $\mathcal{A} - \boldsymbol{M}$ daraus entsteht, d. h. jedes $\mathcal{A}_n$; also auch alles, was durch Limesbildung an auf- oder absteigenden Folgen daraus entsteht, d. h. jede $\mathcal{A}$-Menge (bis auf eine Menge vom Maße 0, vgl. 3. und a. O. Anm. [45]). Dabei ist das Maß der $\mathcal{A}$-Menge, d. h. $\tilde{\mu}^{(0)}(X) = \mu(\boldsymbol{M}[X])$ endlich. Also: alle $\boldsymbol{M}[X]$ mit meßbarem X und endlichem $\tilde{\mu}^{(0)}(X) = \mu(\boldsymbol{M}[X])$ sind, bis auf Mengen vom Maße 0, mit allen $\mathcal{A}$-Mengen identisch.

6. Betrachten wir die Abbildung $[P, t] \to [S_t P, t]$. Ihr Definitionsbereich ist derjenige von $S_t P$, also meßbar; für jedes feste t bilden die P, deren $[P, t]$ nicht in ihm liegt, eine Menge vom Maße 0. Seine Komplementärmenge hat also, nach dem Satz von Fubini, auch als $[P, t]$-Menge das Maß 0. Betrachten wir das Urbild einer Menge $\overline{N}$ (d. i. die Menge der $[P, t]$ mit $[S_t P, t]$ in $\overline{N}$, $\overline{N}$ braucht keineswegs aus lauter $[S_t P, t]$-s zu bestehen!). Ist $\overline{N} = O \times \tilde{O}$, O, $\tilde{O}$ Borelsch, so ist dieses Urbild der Durchschnitt des $[P, t] \to [S_t P]$-Urbildes von O, das wegen der Meßbarkeit von $S_t P$ meßbar ist, mit der offenbar meßbaren Menge der $[P, t]$ mit t in $\tilde{O}$ — also meßbar. Bei gegebenem t ist die Menge der P mit $[P, t]$ im Urbild leer, wenn t nicht zu $\tilde{O}$ gehört, und $S_t^{-1} O$, wenn t zu $\tilde{O}$ gehört. Nach dem Satz von Fubini ist also das Maß des Urbildes

legen, die bei $\xi \to$ ins zu $[x, \xi]$ gehörige P aus der gewöhnlichen Topologie von $0 \leqq \xi \leqq 1$ entsteht. Eine andere denkbare Topologie für $M(x)$ ist die, bei der die in $M(x)$ relativ-offenen Mengen (im Sinne der Ω-Topologie, also die $O \cdot M(x)$, O offen in Ω) als offen gelten. Nun sind die Borelschen Mengen der ersteren Topologie aus den Bildern der ξ-Mengen $\xi \leqq \eta$, η in A' erzeugbar, d. h. aus den Mengen $M(x) \cdot K(\eta)$, oder, was nach Konstruktion der $K(\eta)$ dasselbe ist, aus den $M(x) \cdot K_n$. Die Borelschen Mengen der letzteren Topologie entstehen aus den $M(x) \cdot O$, also auch aus den $M(x) \cdot K_n$. Beide Topologien haben also dieselben Borelschen Mengen, da aber für den l-Maßcharakter nur die Borelschen Mengen maßgebend sind (vgl. Anm. [9] in der vorhergehenden Abhandlung des Verfassers), ist $\mu_{(x)}$ auch im Sinne der letzteren Topologie, d. i. der auf $M(x)$ übertragenen Topologie von Ω, ein l-Maß.

Ist $M[X] = \sum_{x \text{ in } X} M(x)$, so ist $M[X]$ das Bild von $\bar{X}$, also sind auch X und $M[X]$ gleichzeitig meßbar, und haben dieselben ($\tilde{\mu}^{(0)}$- bzw. μ-)Maße. Aus unserer $\bar{\mu}(\bar{C})$-Formel aber wird: für jedes meßbare $N \subseteqq \Omega_1$ ist

$$\mu(N) = \int_0^1 \mu_{(x)}(M(x) \cdot N)\, dw(x).$$

Dabei ist $w(x) = w(x, 1) = \tilde{\mu}^{(0)}$ (Menge aller y mit $0 < y \leqq x$). Der Integrand kann für einige x sinnlos sein ($\mu^*_{(x)}$ undefiniert oder $M(x) \cdot N$ unmeßbar), aber die Menge dieser x hat das $\tilde{\mu}^{(0)}$-Maß 0; insbesondere ist $\mu^*_{(x)}$ nur für die x von X_0 definiert, aber diese Komplementärmenge von X_0 in $0 \leqq x \leqq 1$, Y_0, hat das $\tilde{\mu}^{(0)}$-Maß 0.

Untersuchen wir nunmehr das Verhältnis der S_t zu den $M(x)$. Sei T der m-Raum der t, $\tilde{\mu}$ das gewöhnliche Lebesguesche Maß in ihm, $\Omega \times T$ der Produktraum, $\bar{\mu}$ das dort nach D_1 definierte l-Maß (die Verwechslung mit dem $\bar{\mu}$ in 4. droht nicht, da dieses nicht mehr vorkommen wird).

Sei $M \subseteqq \Omega$. Wenn es offen ist, ist sein $[P, t] \to S_t P$-Urbild in $\Omega \times T$ meßbar (d. h. die Funktion $S_t P$ ist meßbar, vgl. Anm. [11]), und von den offenen M überträgt sich dies mühelos auf die Borelschen. Hat M außerdem das Maß 0, so ist für jedes feste t die Menge der P mit $S_t P$ in M vom Maße 0 (S_t ist maßtreu), also ergibt der Satz von Fubini (a. a. O. Anm. [10]), den wir auf das Urbild wegen seiner Meßbarkeit anwenden dürfen: das Urbild selbst hat das Maß 0. Hat M bloß das Maß 0, so ist $M \subseteqq M'$, M' Borelsch und vom Maß 0, also Urbild von $M \subseteqq$ Urbild von M', das vom Maße 0 ist, also das Urbild von M auch. Das Urbild von M ist also meßbar, wenn M Borelsch ist, oder vom Maße 0, also auch für die Summen zweier solcher M, d. h. für jedes meßbare M.

Gehöre x zu A'. $A(x)$ ist meßbar, also auch die Menge der $[P, t]$, P in $A(x)$, ferner das $S_t P$-Urbild: die Menge der $[P, t]$, $S_t P$ in $A(x)$.

Also ist auch die Summe minus der Durchschnitt dieser zwei Mengen meßbar, d. i. die Menge der $[P, t]$, für die P zu $\mathcal{A}(x)$ gehört, aber $S_t P$ nicht, oder umgekehrt. Und auch die Summe dieser Mengen für alle x von A' ist meßbar, sie heiße $\overline{Y}_1$. Nun sei t fest. Die Menge der P, für die $[P, t]$ zu $\overline{Y}_1$ gehört, ist

$$\sum_{x \text{ in } A'} [(\mathcal{A}(x) + S_t^{-1}\mathcal{A}(x)) - (\mathcal{A}(x) \cdot S_t^{-1}\mathcal{A}(x))],$$

also, wie jeder ihrer Addenden, vom Maße 0. Der Satz von Fubini ist anwendbar: $\overline{Y}_1$ hat das Maß 0. $[P, t]$ nicht in $\overline{Y}_1$ bedeutet: P gehört zu denselben $\mathcal{A}(x)$ (x in A'), wie $S_t P$, also auch zu denselben $\boldsymbol{M}(x)$ ($0 \leqq x < 1$), oder: P gehört dann und nur dann zu $\boldsymbol{M}(x)$, wenn $S_t P$ es tut, u. zw. für alle $0 \leqq x < 1$.

Nun definieren wir $S_t P$ um, indem wir es in $\overline{Y}_1$ für undefiniert erklären, sonst aber nicht ändern. Da $\overline{Y}_1$ das Maß 0 hat, bleibt $S_t P$ meßbar; da bei festem t die Menge der P mit $[P, t]$ in $\overline{Y}_1$ das Maß 0 hat, sind die neuen S_t den alten äquivalent. Dabei sind alle $\boldsymbol{M}(x)$ den neuen S_t gegenüber genau invariant. Auf Grund der Ersetzungsmöglichkeit durch eine äquivalente allgemeine Strömung dürfen wir also annehmen, daß die $\boldsymbol{M}(x)$ von vornherein genau invariant waren.

Da jedes $\boldsymbol{M}(x)$ genau S_t-invariant ist, ist es auch jedes $\boldsymbol{M}[X]$. Ist X meßbar, $\tilde{\mu}^{(0)}(X)$ endlich, so ist auch $\boldsymbol{M}[X]$ meßbar, $\mu(\boldsymbol{M}[X])$ endlich, d. h. $\boldsymbol{M}[X]$ $\mathcal{A}$-Menge. Umgekehrt ist jedes $\mathcal{A}(\eta)$ ein $\boldsymbol{M}[X]$ (X ist die Menge $0 \leqq \xi \leqq \eta$) mit meßbarem X; also auch alles, was durch $\mathcal{A} + \boldsymbol{M}$, $\mathcal{A} \cdot \boldsymbol{M}$, $\mathcal{A} - \boldsymbol{M}$ daraus entsteht, d. h. jedes $\mathcal{A}_n$; also auch alles, was durch Limesbildung an auf- oder absteigenden Folgen daraus entsteht, d. h. jede $\mathcal{A}$-Menge (bis auf eine Menge vom Maße 0, vgl. 3. und a. O. Anm. [45]). Dabei ist das Maß der $\mathcal{A}$-Menge, d. h. $\tilde{\mu}^{(0)}(X) = \mu(\boldsymbol{M}[X])$ endlich. Also: alle $\boldsymbol{M}[X]$ mit meßbarem X und endlichem $\tilde{\mu}^{(0)}(X) = \mu(\boldsymbol{M}[X])$ sind, bis auf Mengen vom Maße 0, mit allen $\mathcal{A}$-Mengen identisch.

6. Betrachten wir die Abbildung $[P, t] \to [S_t P, t]$. Ihr Definitionsbereich ist derjenige von $S_t P$, also meßbar; für jedes feste t bilden die P, deren $[P, t]$ nicht in ihm liegt, eine Menge vom Maße 0. Seine Komplementärmenge hat also, nach dem Satz von Fubini, auch als $[P, t]$-Menge das Maß 0. Betrachten wir das Urbild einer Menge $\overline{N}$ (d. i. die Menge der $[P, t]$ mit $[S_t P, t]$ in $\overline{N}$, $\overline{N}$ braucht keineswegs aus lauter $[S_t P, t]$-s zu bestehen!). Ist $\overline{N} = O \times \tilde{O}$, O, $\tilde{O}$ Borelsch, so ist dieses Urbild der Durchschnitt des $[P, t] \to [S_t P]$-Urbildes von O, das wegen der Meßbarkeit von $S_t P$ meßbar ist, mit der offenbar meßbaren Menge der $[P, t]$ mit t in $\tilde{O}$ — also meßbar. Bei gegebenem t ist die Menge der P mit $[P, t]$ im Urbild leer, wenn t nicht zu $\tilde{O}$ gehört, und $S_t^{-1} O$, wenn t zu $\tilde{O}$ gehört. Nach dem Satz von Fubini ist also das Maß des Urbildes

$$\int_{\tilde{O}} \mu(S_t^{-1} O)\, dt = \int_{\tilde{O}} \mu(O)\, dt = \mu(O) \cdot \tilde{\mu}(\tilde{O})$$
$$= \bar{\mu}(O \times \tilde{O}) = \bar{\mu}(\bar{N}) \text{ (vgl. } D_1).$$

Die Aussage: das Urbild von $\bar{N}$ ist meßbar und hat dasselbe Maß wie $\bar{N}$, gilt also für $\bar{N} = O \times \tilde{O}$, O, $\tilde{O}$ Borelsch. Somit gilt sie auch für alle endlichen $O \times \tilde{O}$-Summen (da die Addenden fremd wählbar sind), für Limites aufsteigender Folgen derselben, d. i. insbesondere für alle offenen Mengen. Und weiter überträgt sie sich durch Limesbildung an auf- oder absteigenden Folgen auf alle Borelschen $\bar{N}$. Da jedes $\bar{N}$ vom Maße 0 Teilmenge eines Borelschen $\bar{N}'$ vom Maße 0 ist, gilt sie auch für diese, also auch für die Summen Borelscher Mengen und solcher vom Maße 0—d. h. für alle meßbaren $\bar{N}$.

$[P, t] \to [S_t, P]$ bildet also $\Omega \times T$ minus eine Menge vom Maße 0 eineindeutig auf eine gewisse Teilmenge $\bar{N}^{(0)}$ von $\Omega \times T$ ab, und das Urbild jedes meßbaren $\bar{N}$ ist meßbar und vom selben Maße wie $\bar{N}$ (wobei $\bar{N}$ keineswegs $\subseteq \bar{N}^{(0)}$ zu sein braucht). Betrachtet man den Beweis von Satz 2 in der vorhergehenden Abhandlung des Verfassers, so erkennt man, daß er auf diesen Fall anwendbar ist, und $\bar{N}^{(0)} = \Omega \times T$ minus eine Menge vom Maße 0 ergibt. Ist dies gesichert, so kann man den genannten Satz 2 selbst anwenden, aus ihm folgt dann die Maßtreue von $[P, t] \to [S_t P, t]$. Die inverse Abbildung $[P, t] \to [S_t^{-1} P, t]$ ist also auch maßtreu.

Sei N meßbar. Ist $\bar{N}$ die Menge aller $[P, t]$, P in N, so ist es auch meßbar. Somit ist sein $[S_t P, t]$- und sein $[S_t^{-1} P, t]$-Urbild meßbar, d. h. das $[P, t] \to S_t P$- und das $[P, t] \to S_t^{-1} P$-Urbild von N (aus der bloßen Meßbarkeit von $S_t P$ würde nur das erstere, und nur für offene oder höchstens für Borelsche N folgen). Also ist $S_t^{-1} P$ meßbar als $[P, t]$-Funktion.

Wir können also $S_t P$ umdefinieren: für $t > 0$ sei es das alte, für $t = 0$ die Identität, für $t < 0$ $S_{-t}^{-1} P$. Das neue $S_t P$ ist also meßbar (in $[P, t]$), und nach D_2' dem alten äquivalent, es ist aber jetzt genau (nicht bloß bis auf Mengen vom Maße 0) $S_t S_{-t} = S_{-t} S_t = S_0 =$ Identität. Wir dürfen also annehmen, daß dies von vornherein so war (die Invarianz der $M(x)$ wird dadurch nicht berührt).

Betrachten wir ferner das $S_t S_s P$-Urbild eines meßbaren N. Zunächst bei festem t: dies ist das $S_s P$-Urbild von $S_t^{-1} N$, welches meßbar ist, also meßbar. Also ist $S_t S_s P$ als $[P, s]$-Funktion meßbar. Zweitens sei alles variabel. Dann bilden wir zuerst das $[P, t] \to S_t P$-Urbild von N, $\bar{N}'$, welches meßbar ist, und sodann das $[P, t, s] \to [S_s P, t]$-Urbild von $\bar{N}'$, $\bar{N}''$, das auch meßbar ist. Letzteres kann genau so bewiesen werden, wie weiter oben das Entsprechende für das $[P, t] \to [S_t P]$-Urbild, nur ist an Stelle der maßtreuen Abbildung $P \to S_t P$ (t fest) die ebenfalls maßtreue

Abbildung $[P, t] \to [S_s P, t]$ (s fest) als Ausgangspunkt zu nehmen. Das meßbare $\bar{N}''$ ist das $[P, t, s] \to S_t S_s P$-Urbild von N. Also ist $S_t S_s P$ auch als $[P, t, s]$-Funktion meßbar.

Schließlich wollen wir die Formel $\mu(N) = \int_0^1 \mu_{(x)} (M(x) \cdot N)\, dw(x)$ ($N \subseteqq \Omega_1$) auf $\Omega_1 \times T$ und $\Omega_1 \times T \times T$ übertragen. Wenn ein $\bar{N} \subseteqq \Omega_1 \times T$ oder ein $\bar{\bar{N}} \subseteqq \Omega_1 \times T \times T$ gegeben ist, verstehen wir unter $(\bar{N})_{(t)}$ bzw. $(\bar{\bar{N}})_{(t,s)}$ die Menge aller P mit $[P, t]$ in $\bar{N}$ bzw. $[P, t, s]$ in $\bar{\bar{N}}$ (also $\subseteqq \Omega_1$). Die neuen Formeln lauten:

$$\bar{\mu}(\bar{N}) = \int_0^1 \int_{-\infty}^{\infty} \mu_{(x)} (M(x) \cdot (\bar{N})_{(t)})\, dw(x)\, dt,$$

$$\bar{\bar{u}}(\bar{\bar{N}}) = \int_0^1 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu_{(x)} (M(x) \cdot (\bar{\bar{N}})_{(t,s)})\, dw(x)\, dt\, ds$$

($\bar{N} \subseteqq \Omega_1 \times T$, $\bar{\bar{N}} \subseteqq \Omega_1 \times T \times T$, der Integrand soll bis auf eine $[x, t]$- bzw. $[x, t, s]$-Menge vom Maße 0 sinnvoll sein, und meßbare Funktion von $[x, t]$ bzw. $[x, t, s]$). Für $\bar{N} = O \times \tilde{O}$ bzw. $\bar{\bar{N}} = O \times \tilde{O} \times \tilde{O}'$ (vgl. D_1) folgt dies unmittelbar aus der ursprünglichen Formel, und von diesen überträgt man es, wie beim Maßtreuebeweise w. o., sukzessiv auf alle offenen, Borelschen, vom Maße 0, und alle meßbaren Mengen $\bar{N}$ bzw. $\bar{\bar{N}}$.

7. Die Menge $\bar{\bar{N}}_2$ der $[P, t, s]$ mit $S_t S_s P \neq S_{t+s} P$ ist meßbar, weil auf beiden Seiten meßbare Funktionen stehen. Bei festem t, s ist die P-Menge mit zu ihr gehörigem $[P, t, s]$ vom Maße 0 (D_2'), also ist sie es selbst auch. Somit ist

$$\int_0^1 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu_{(x)} (M(x) \cdot (\bar{\bar{N}}_2)_{(t,s)})\, dw(x)\, dt\, ds = 0,$$

$$\text{d. h.} \quad \mu_{(x)} (M(x) \cdot (\bar{\bar{N}}_2)_{(t,s)}) = 0,$$

abgesehen von einer $[x, t, s]$-Menge, $\bar{\bar{Y}}_2$, vom Maße 0. Die $\mu_{(x)}$-Bedingung besagt: $S_t S_s P = S_{s+t} P$ für alle P von $M(x)$, bis auf eine Menge vom ($\mu_{(x)}$-) Maß 0, oder: $S_t S_s \cong S_{t+s}$ in $M(x)$ (D_2'). Hätten wir t von vornherein festgehalten, und wären wir von der Meßbarkeit der $[P, s]$-Menge mit $S_t S_s P \neq S_{t+s} P$ ausgegangen, so hätten wir ebenso erhalten: abgesehen von einer $[x, s]$-Menge vom Maße 0, gilt $S_t S_s \cong S_{t+s}$ in $M(x)$, d. h. $[x, t, s]$ gehört nicht zu $\bar{\bar{Y}}_2$.

Wenden wir nun den Satz von Fubini auf $\bar{\bar{Y}}_2$ an: Abgesehen von einer $[x, t]$-Menge vom Maße 0, $\bar{Y}_2$, gehören alle $[x, t, s]$, bis auf eine s-Menge vom Maße 0, nicht zu $\bar{\bar{Y}}_2$. Dasselbe gilt wegen des zuletzt Gesagten auch bei festem t, für alle x bis auf eine Menge vom Maße 0. D. h.: $\bar{Y}_2$ hat das Maß 0, und bei festem t hat die x-Menge mit $[x, t]$ in $\bar{Y}_2$ auch das Maß 0.

Sei N meßbar, $\overline{N}$ sein $[P, t] \to S_t P$-Urbild, also auch meßbar.

$$\mu_{(x)}(M(x) \cdot (\overline{N})_{(t)}) = \mu_{(x)}(M(x) \cdot S_t^{-1} N)$$

ist dann eine meßbare $[x, t]$-Funktion, ebenso $\mu_{(x)}(M(x) \cdot N)$. Die $[x, t]$-Menge $\overline{Y}_3(N)$ der $\mu_{(x)}(M(x) \cdot S_t^{-1} N) \neq \mu_{(x)}(M(x) \cdot N)$ ist also meßbar. Wir wenden nun unsere μ-, $\mu_{(x)}$-Formel auf $M[X] \cdot S_t^{-1} N$ und $M[X] \cdot N$ an, es ist:

$$\begin{aligned} \mu(M[X] \cdot S_t^{-1} N) &= \int_0^1 \mu_{(x)}(M(x) \cdot M[X] \cdot S_t^{-1} N)\, dw(x) \\ &= \int_X \mu_{(x)}(M(x) \cdot S_t^{-1} N)\, dw(x), \\ \mu(M[X] \cdot N) &= \int_0^1 \mu_{(x)}(M(x) \cdot M[X] \cdot N)\, dw(x) \\ &= \int_X \mu_{(x)}(M(x) \cdot N)\, dw(x). \end{aligned}$$

Wegen der Maßtreue von S_t und der Invarianz der $M(x)$, $M[X]$ sind die linken Seiten gleich, also auch die rechten. Dies gilt für alle X, daher ist bis auf eine x-Menge vom Maße 0

$$\mu_{(x)}(M(x) \cdot S_t^{-1} N) = \mu_{(x)}(M(x) \cdot N).$$

D. h.: für jedes feste t ist die x-Menge mit $[x, t]$ in $\overline{Y}_3(N)$ vom Maße 0. Nach dem Satz von Fubini ist daher auch $\overline{Y}_3(N)$ vom Maße 0.

Sei nun $\overline{Y}_3 = \sum_{\xi \text{ in } A'} \overline{Y}_3(K(\xi))$. Auch für $\overline{Y}_3$ gilt: es hat das Maß 0, und bei festem t hat die x-Menge mit $[x, t]$ in $\overline{Y}_3$ auch das Maß 0. Liegt $[y, t]$ nicht in $\overline{Y}_3$, so gilt

$$\mu_{(y)}(S_t^{-1} M) = \mu_{(y)}(M) \qquad (M, S_t^{-1} M \text{ meßbar!})$$

für alle $M = M(y) \cdot K(\eta)$, d. h. für diejenigen Urbilder $M \subseteqq M(y)$, die nach der Abbildung $P \to [x, \xi]$ aus 3 den Strecken $[x, \xi]$ mit $x = y$, $0 \leqq \xi \leqq \eta$ entsprechen. Hieraus folgt, daß es auch für die Urbilder von Differenzen solcher Strecken gilt, sowie für die von Summen endlich vieler solcher Differenzen und für die Limites aufsteigender Folgen solcher Summen. D. i. für die Urbilder von offenen Mengen, von welchen man auf die gewohnte Weise zu den Urbildern Borelscher, vom Maße 0 ($\tilde{\mu}_{(y)}$-Maß!, diese Mengen haben wegen der maßtreuen Übertragung selbst das $\mu_{(y)}$-Maß 0), und meßbarer Mengen gelangt. Diese letzteren Urbilder sind aber einfach alle ($\mu_{(y)}$-) meßbaren $M \subseteqq M(y)$, unsere Formel gilt mithin auch für diese. Für $M = M(y)$ insbesondere wird demnach

$$\mu_{(y)}(S_t^{-1} M(y)) = \mu_{(y)}(M(y)),$$

und da $S_t^{-1} M(y) \subseteqq M(y)$ und $\mu_{(y)}(M(y))$ endlich ist (es ist $\leqq 1$, denn $M(y)$ ist der Limes der Urbilder der $\xi < \zeta \leqq \eta$, $\xi \to 0$, $\zeta \to 1$ (vgl. 3), und diese haben die Maße $\varrho(y, \eta) - \varrho(y, \xi) \leqq 1$), ist $M(y) - S_t^{-1} M(y)$ vom Maße 0. D. h. $S_{-t} = S_t^{-1}$ ist überall in $M(y)$ definiert bis auf eine Menge vom ($\mu_{(x)}$-) Maße 0.

Wenn also $[x, t]$, $[x, -t]$ beide nicht in $\overline{Y}_3$ liegen, so bildet $S_t\ M(x)$ minus eine Menge vom ($\mu_{(x)}$-) Maße 0 auf $M(x)$ minus eine Menge vom ($\mu_{(x)}$-) Maße 0 ab und ist ($\mu_{(x)}$-) maßtreu.

8. Die Menge der $[x, t]$, die in $\overline{Y}_2$ oder $\overline{Y}_3$ liegen oder für die $[x, -t]$ in $\overline{Y}_3$ liegt, heiße $\overline{Y}_4$. $\overline{Y}_4$ hat wieder das Maß 0, und bei festem t hat die Menge der x mit $[x, t]$ in $\overline{Y}_4$, d. i. $(\overline{Y}_4)_{(t)}$, das Maß 0. Gehört $[x, t]$ nicht zu $\overline{Y}_4$, so bildet $S_t\ M(x)$ bis auf Mengen vom Maße 0 (vgl. am Ende von 7) eineindeutig und maßtreu auf sich selbst ab, und für alle s, bis auf eine Menge vom Maße 0, gilt $S_t S_s \cong S_{t+s}$ in $M(x)$.

Wenn $\overline{X} \subseteqq A \times T$ ist, so sei $\overline{M}[\overline{X}]$ die Menge aller $[P, t]$, für die P zu einem $M(x)$ mit $[x, t]$ in $\overline{X}$ gehört. So wie wir am Ende von 4 zeigten: mit X ist auch $M[X]$ meßbar und vom selben Maß ($\tilde{\mu}^{(0)}$ bzw. μ), gilt auch: mit $\overline{X}$ ist auch $\overline{M}[\overline{X}]$ meßbar und vom selben Maß ($\tilde{\overline{\mu}}^{(0)}$ bzw. $\overline{\mu}$). Denn für $\overline{X} = X \times \tilde{O}$ (vgl. D_1) folgt dies, aus der Behauptung über die $X, M[X]$, und von diesen überträgt es sich auf die gewohnte Weise sukzessiv auf die offenen, Borelschen, vom Maße 0, und die meßbaren $\overline{X}$.

Also hat $\overline{M}[\overline{Y}_4]$ das ($\overline{\mu}$-) Maß 0, und bei festem t hat $M[(\overline{Y}_4)_{(t)}]$, welches offenbar mit der Menge der P mit $[P, t]$ in $\overline{M}[Y_4]$ zusammenfällt, ebenfalls das (μ-) Maß 0. Wenn wir also die Definition von $S_t P$ innerhalb von $\overline{M}[\overline{Y}_4]$ beliebig ändern, wird das neue $S_t P$ allenfalls eine meßbare $[P, t]$-Funktion, und es bleibt dem alten äquivalent (vgl. D_2'), wir erhalten also eine äquivalente allgemeine Strömung. Der zulässige Änderungsbereich läßt sich auch so kennzeichnen: S_t wird innerhalb von $M(x)$ geändert, falls $[x, t]$ zu $\overline{Y}_4$ gehört.

Wenn $\overline{N}$ meßbar und $\subseteqq \Omega_1 \times T$ ist, so gilt

$$\overline{\mu}(\overline{N}) = \int_0^1 \overline{\mu}_{(x)}((M(x) \times T) \cdot \overline{N})\, dw(x)$$

($\overline{\mu}_{(x)}$ das nach D_1 von $M(x)$ auf $M(x) \times T$ erweiterte $\mu_{(x)}$, der Integrand ist, bis auf eine x-Menge vom Maße 0 sinnvoll, d. h. $(M(x) \times T) \cdot N$ $\overline{\mu}_{(x)}$-meßbar). Dies folgt für $\overline{N} = O \times \tilde{O}$ (vgl. D_1) aus unserer ursprünglichen μ-, $\mu_{(x)}$-Formel und wird auf die übrigen $\overline{N}$ so ausgedehnt, wie es für die analogen Formeln am Ende von 6 geschah.

Sei nun N meßbar, $\subseteqq \Omega_1$, und $\bar{N}$ sein $[P, t] \to S_t P$-Urbild. Bis auf eine x-Menge vom Maße 0, $Y_5(N)$, ist $(M(x) \times T) \cdot \bar{N}$ ($\bar{\mu}_{(x)}$-) meßbar, d. i. das $[P, t] \to [S_t P]$-Urbild von $M(x) \cdot N$. Sei $Y_5 = \sum_{\xi \text{ in } A'} Y_5(K(\xi))$, auch dieses hat das Maß 0. Liegt y nicht in Y_5, so sind die $[P, t] \to S_t P$-Urbilder aller $M(y) \cdot K(\eta)$ ($\bar{\mu}_{(y)}$-) meßbar, d. h. die Urbilder jener $M \subseteqq M(y)$, die vermöge der $P \to [x, \xi]$-Abbildung von 3. den Intervallen $x = y$, $0 \leqq \xi \leqq \eta$ entsprechen. Dasselbe gilt dann auch für diejenigen M, die Differenzen solcher Intervalle entsprechen, oder Summen endlich vieler solcher Differenzen, oder Limites aufsteigender Folgen solcher Summen — d. h. allen offenen Mengen auf $x = y$. Somit ist $S_t P$ als $[P, t]$-Funktion in $M(x) \times T$ ($\bar{\mu}_{(x)}$-) meßbar.

Für jedes x möge die Menge der t mit $[x, t]$ in $\bar{Y}_4$ $T(x)$ heißen. Nach dem Satz von Fubini ist $T(x)$ stets eine t-Menge vom Maße 0, abgesehen von einer x-Menge vom Maße 0, Y_4.

Die Summe von Y_0, Y_4, Y_5 heiße Y_6, dies ist auch eine Menge vom Maße 0, ihre Komplementärmenge (in $0 \leqq x \leqq 1$) heiße X_6. Für die x von X_6 gilt folgendes: $\mu_{(x)}$ ist definiert, und ein l-Maß in $M(x)$. $T(x)$ hat das Maß 0; liegt t nicht in $T(x)$, so bildet S_t $M(x)$ minus eine Menge vom Maße 0 eineindeutig und maßtreu auf $M(x)$ minus eine Menge vom Maße 0 ab, für alle s bis auf eine Menge vom Maße 0 gilt $S_t S_s \cong S_{t+s}$ in $M(x)$ (wir können, ohne hieran zu ändern, auch s, $t+s$ als nicht in $T(x)$ liegend annehmen), $S_t P$ ist eine meßbare $[P, t]$-Funktion (in $M(x)$ bzw. $M(x) \times T$, gemeint sind hier stets die Maße $\mu_{(x)}$ bzw. $\bar{\mu}_{(x)}$). Im nächsten Paragraphen werden wir zeigen, daß unter diesen Umständen die S_t der t in $T(x)$ so umdefiniert bzw. neudefiniert werden können, daß diese S_t zusammen D_2' in $M(x)$ (mit $\mu_{(x)}$) erfüllen, d. h. dort eine allgemeine Strömung bilden. Da für t von $T(x)$ $[x, t]$ zu $\bar{Y}_4$ gehört, bleibt nach dem früher Gesagten die neue S_t-Strömung der alten äquivalent. Wir dürfen also annehmen, daß das soeben Erreichte von vornherein der Fall war: daß für jedes x von X_6 die S_t eine allgemeine Strömung in $M(x)$ (mit $\mu_{(x)}$) bilden.

9. In 3. definierten wir: $\tilde{\mu}^{(0)}(X) = \int_X d\bar{\mu}(x, 1) = \int_X dw(x)$, und in 4. zeigten wir: $\tilde{\mu}^{(0)}(X) = \mu(M[X])$, was nach der μ-, $\mu_{(x)}$-Formel gleich

$$\int_0^1 \mu_{(x)}(M(x) \cdot M[X])\, dw(x) = \int_X \mu_{(x)}(M(x))\, dw(x)$$

ist. Somit gilt allgemein

$$\int_X dw(x) = \int_X \mu_{(x)}(M(x))\, dw(x),$$

es muß also, bis auf eine Menge Y_7 vom Maße 0, $\mu_{(x)}(M(x)) = 1$ sein.

Wir bilden zu jedem $M(x)$ den Hilbertschen Raum $\mathfrak{H}_{(x)}$ der in ihm definierten Funktionen $f_{(x)}(P)$ mit endlichem $\int_{M(x)} |f_{(x)}(P)|^2 \, d v_{(x,P)}$ (das Integral ist auf Grund des l-Maßes $\mu_{(x)}$ zu bilden, $v_{(x,P)}$ ist das entsprechende Volumelement). Da $P \to S_t P$ in $M(x)$ maßtreu ist, können wir die unitären Koopmanschen Operatoren $U_{(x,t)}$ bilden, und das dazugehörige $E_{(x,0)}$ (vgl. 1.). Ferner sei $\chi_{(x,\xi)}(P) \begin{cases} = 1 \text{ in } M(x) \cdot K(\xi) \\ = 0 \text{ sonst in } M(x) \end{cases}$.

Ist $f_{(x)}(P)$ zu allen $\chi_{(x,\xi)}(P)$ (ξ in A') orthogonal, so wenden wir die Abbildung $P \to [y, \eta]$ aus 3. an. Dabei wird $y = x$, $0 \leqq \eta \leqq 1$. Aus $f_{(x)}(P)$ wird also ein $f'_{(x)}(\eta)$, aus $\chi_{(x,\xi)}(P)$ aber $\chi'_{(x,\xi)}(\eta) \begin{cases} = 1 \text{ für } 0 \leqq \eta \leqq \xi \\ = 0 \text{ sonst} \end{cases}$. $f'_{(x)}$, $\chi'_{(x,\xi)}$ sind wieder orthogonal, d. h. $\int_0^\xi f'_{(x)}(\eta) \, d_{(x)} \eta = 0$ ($d_{(x)}$ deute an, daß das Integral auf Grund des $\tilde{\mu}_{(x)}$-Maßes zu bilden ist) für alle ξ. Also ist $f'_{(x)}(\eta) = 0$, bis auf eine η-Menge vom ($\tilde{\mu}_{(0)}$-) Maße 0, d. h. $f_{(x)}(P) = 0$, bis auf eine P-Menge vom ($\mu_{(x)}$-) Maße 0. Da somit nur 0 zu allen $\chi_{(x,\xi)}$ (ξ in A') orthogonal ist, spannen diese die abgeschlossene Linearmannigfaltigkeit $\mathfrak{H}_{(x)}$ auf.

Wir setzen $\xi_0 = 1$, $\xi_1, \xi_2, \cdots$ irgendeine Abzählung von A', und orthogonalisieren die $\chi_{(x,\xi_n)}$, $n = 0, 1, 2, \cdots$, in dieser Reihenfolge[53]. So entstehen Funktionen $\varphi_{(x,m)} = \sum_{n=0}^{\infty} c_{(x,m,n)} \, \chi_{(x,\xi_n)}$, $m = 0, 1, 2, \cdots$, die nach der obigen Bemerkung ein vollständiges normiertes Orthogonalsystem bilden. Da $\chi_{(x,\xi_0)} = 1$ (für alle x in $M(x)$) normiert ist, und der Prozeß mit ihm beginnt, ist $\varphi_{(x,0)} = \chi_{(x,\xi_0)} = 1$. Die $c_{(x,m,n)}$ entstehen durch elementare Rechenoperationen aus den

$$(\chi_{(x,\xi_m)}, \chi_{(x,\xi_n)}) = \mu_{(x)} (M(x) \cdot A(\xi_m) \cdot A(\xi_n)),$$

da diese meßbare x-Funktionen sind, sind es auch die $c_{(x,m,n)}$.

Sei t fest.

$$(S_t \chi_{(x,\xi_m)}, \chi_{(x,\xi_n)}) = \mu_{(x)} (M(x) \cdot S_t^{-1} A(\xi_m) \cdot A(\xi_n))$$

ist eine meßbare x-Funktion, da es die $c_{(x,m,n)}$ auch sind, sind es die $\varphi_{(x,m)}$-Matrizenelemente der S_t, $s_{(t,x,m,n)} = (S_t \varphi_{(x,m)}, \varphi_{(x,n)})$ gleichfalls. Da nach 1

$$E_0 = \operatorname*{limes}_{t \to \infty} \frac{1}{s-t} \int_t^s S_\tau \, d\tau = \operatorname*{limes}_{p \to \infty} \left(\operatorname*{limes}_{q \to \infty} \frac{1}{2pq} \sum_{\nu=-pq}^{+pq} S_{\frac{\nu}{q}} \right)$$

$$(p, q = 1, 2, \cdots, \nu = 0, \pm 1, \pm 2, \cdots)$$

[53] Nach dem bekannten Verfahren von E. Schmidt, vgl. auch E S. 69.

ist[54], sind auch die $\varphi_{(x,m)}$-Matrizenelemente der E_0, $e_{(0,x,m,n)} = (E_0\,\varphi_{(x,m)}, \varphi_{(x,n)})$ meßbare x-Funktionen.

Da $\varphi_{(x,0)}$ konstant ist, ist $E_0\,\varphi_{(x,0)} = \varphi_{(x,0)}$, also da E_0 Projektionsoperator ist, $e_{(0,x,m,n)} \begin{cases} = 1, \text{ für } m = n = 0 \\ = 0, \text{ für } m = 0,\ n \neq 0 \\ = 0, \text{ für } m \neq 0,\ n = 0 \end{cases}$. Daß S_t in $M(x)$ ergodisch ist, besagt, daß $E_0 f_{(x)} = f_{(x)}$ keine Lösungen außer $f_{(x)} = c\,\varphi_{(x,0)}$ hat, d. h. $e_{(0,x,m,n)} = 0$, für $m, n \neq 0$. Die Menge der x, für die das nicht gilt, Y_8, ist somit meßbar. Sei ein m gegeben, die x-Menge für die es bei diesem m und mindestens einem n nicht gilt, $Y_{8,m}$ ist ebenfalls meßbar, und es ist $Y_8 = \sum_{m=1}^{\infty} Y_{8,m}$. Wir werden zeigen, daß alle $Y_{8,m}$ das Maß 0 haben, also auch Y_8. Nehmen wir also an, ein $Y_{8,m}$ hätte nicht das Maß 0.

Gehöre x zu $Y_{8,m}$. Da E_0 Projektionsoperator ist, ist $f_{(x)} = \sum_{0}^{\infty n} e_{(0,x,m,n)}\,\varphi_{(x,n)}$ Lösung von $E_0 f_{(x)} = f_{(x)}$, also von $S_t f_{(x)} = f_{(x)}$, und nach dem was wir schon wissen, orthogonal zu $\varphi_{(x,0)}$ = Konstans und $\neq 0$ (auch nicht bis auf Mengen vom Maße 0!). $f_{(x)}(P)$ ist aus den $e_{(0,x,m,n)}$, $c_{(x,m,n)}$, $\chi_{(x,\xi_m)}(P)$ aufgebaut. Lassen wir P über Ω_1 variieren, und x durch die Bedingung P in $M(x)$ bestimmen, so ist $\chi_{(x,\xi_m)}(P) \begin{cases} = 1, \text{ für } P \text{ in } K(\xi_m) \\ = 0, \text{ sonst} \end{cases}$, also ($\mu$-) meßbar in P. $e_{(0,x,m,n)}$, $c_{(x,m,n)}$ sind ($\tilde{\mu}^{(0)}$-) meßbar in x, also auch (μ-) meßbar in P (denn jede meßbare x-Menge X hat als P-Urbild $M[X]$, das auch meßbar ist). Also ist $f_{(x)}(P)$ auch als P-Funktion (μ-) meßbar. Sei C die Menge der komplexen Zahlen $u + i\,v$ mit $u > 0$ oder $u = 0$, $v > 0$, D die mit $u < 0$ oder $u = 0$, $v < 0$. Dann sind auch die P-Mengen $f_{(x)}(P)$ in C und $f_{(x)}(P)$ in D (μ-) meßbar, also die x-Funktionen

$$\mu_{(x)}\ (P\text{-Menge in } M(x),\quad f_{(x)}(P) \text{ in } C),$$
$$\mu_{(x)}\ (P\text{-Menge in } M(x),\quad f_{(x)}(P) \text{ in } D)$$

beide ($\tilde{\mu}^{(0)}$-) meßbar. Bei festem x ist $f_{(x)}(P)$ (P in $M(x)$) nicht bis auf eine Menge vom Maße 0 gleich 0, und zur Konstanten orthogonal, also muß es in einer Menge vom Maße > 0 in C liegen und in einer ebensolchen Menge in D liegen. D. h. die obigen Funktionen sind stets > 0. Wir können also ein $\varepsilon > 0$ finden, so daß beide in einer Teilmenge Y' von Y_8, die ein Maß > 0 hat, stets $\geqq \varepsilon$ sind. Wir können dabei Y' auch mit endlichem Maße annehmen. Wir nennen nun die Menge der P aus $M[Y']$ mit $f_{(x)}(P)$ in C bzw. D, J bzw. K.

Nach Entstehen sind J, K fremd, beide (μ-) meßbar. Die P-Menge, auf der sich $f_{(x)}(P)$, $f_{(x)}(S_t P)$ unterscheiden, ist (μ-) meßbar, für jedes x ist ihr in $M(x)$ liegender Teil (wegen $U_t f_{(x)} = f_{(x)}$ in $M(x)$) vom ($\mu_{(x)}$-) Maße 0,

[54] Die Limites gelten im Sinne der starken Operatoren-Konvergenz, und wir brauchen nur die Konvergenz der Matrizenelemente, d. h. die schwache. Vgl. a. a. O. Anm. [43].

also hat sie (nach unserer μ-, $\mu_{(x)}$- Formel) selbst das Maß 0. Also sind J, K bis auf Mengen vom (μ-) Maße 0, S_t-invariant. Sie sind ferner $\subseteqq M[Y']$, $\mu(M[Y']) = \tilde{\mu}^{(0)}(Y')$ ist endlich, also sind J, K $\mathfrak{M}$-Mengen. Nach dem am Ende von 5. Gesagten sind sie also, bis auf Mengen vom (μ-) Maße 0, gleich $M[Z]$, $M[W]$.

Da aber in Y' stets $\mu_{(x)}(M(x) \cdot J) \geqq \varepsilon$ ist, gilt für jedes ($\tilde{\mu}^{(0)}$- meßbare) $V \subseteqq Y'$

$$\mu_{(x)}(M(V) \cdot J) \geqq \varepsilon \cdot \tilde{\mu}^{(0)}(V) \quad (\mu\text{-}, \mu_{(x)}\text{- Formel!}),$$

ebenso

$$\mu(M(V) \cdot K) \geqq \varepsilon \cdot \tilde{\mu}^{(0)}(V).$$

Setzen wir in der ersten Formel $V = W$, so wird

$$\varepsilon \cdot \tilde{\mu}^{(0)}(W) \leqq \mu(M(W) \cdot J) = \mu(K \cdot J) = 0,$$
$$\tilde{\mu}^{(0)}(W) = 0, \qquad \mu(K) = \mu(M[W]) = 0,$$

und setzen wir nun in der zweiten Formel $V = Y'$, so wird

$$\varepsilon \cdot \tilde{\mu}^{(0)}(Y') \leqq \mu(M[Y'] \cdot K) = \mu(K) = 0, \qquad \tilde{\mu}^{(0)}(Y') = 0.$$

Aber dies widerspricht der ursprünglichen Annahme über Y'.

Somit muß Y_8 das Maß Null haben.

10. Sei Y_9 die Summe von Y_6, Y_7, Y_8, also auch vom Maße 0, X_9 seine Komplementärmenge (in $0 \leqq x \leqq 1$). Für die x von X_9 ist also $\mu_{(x)}$ ein l-Maß in $M(x)$, S_t ein allgemeiner Fluß in $M(x)$, $\mu_{(x)}(M(x)) = 1$, S_t ergodisch in $M(x)$.

Es ist $\mu(M[Y_9]) = \tilde{\mu}^{(0)}(Y_9) = 0$, wenn wir es also zu Ω_2 schlagen, behält dieses seine Eigenschaften aus 3., und es wird $\mu(M[X_9]) = \Omega_1$. Hat dieses Ω_2 das Maß 0, so können wir es zu einem $M(x)$ schlagen, $\mu_{(x)}(\Omega_2) = 0$ definierend, dann bleibt $M[X_9] = \Omega_1$ erhalten, und es wird $\Omega_2 = 0$. Wir können also annehmen, daß diese Forderungen erfüllt sind, ferner wollen wir X_9 lieber $X^{(0)}$ nennen. Dann können wir zusammenfassend aussagen:

SATZ 2. *Ω kann folgendermaßen zerlegt werden: $\Omega = \Omega_1 + \Omega_2$, $\Omega_1 = \sum_{x \text{ in } X^{(0)}} M(x)$, wobei $X^{(0)}$ eine Menge von Zahlen in $0 \leqq x \leqq 1$ ist, und die $M(x)$ sowie Ω_1, Ω_2 paarweise elementfremde Mengen. $X^{(0)}$ und die $M(x)$ sind m-Räume (indem die gewöhnliche Topologie von $0 \leqq x \leqq 1$ bzw. von Ω auf sie übertragen wird, vgl. Anfang von 5.), in ihnen sind l-Maße $\tilde{\mu}^{(0)}$ bzw. $\mu_{(x)}$ definiert. Diese Zerlegung hat, nachdem evtl. S_t durch eine äquivalente allgemeine Strömung ersetzt wurde, die folgenden Eigenschaften:*

1. *Ω_2 ist entweder leer, oder $\mu(\Omega_2) = \infty$ und Ω_2 null-ergodisch.*
2. *In jedem $M(x)$ ist S_t (unter Zugrundelegung des l-Maßes $\mu_{(x)}$ und der auf $M(x)$ übertragenen relativen Topologie aus Ω) eine eigentlich-ergodische allgemeine Strömung.*

3. *Es ist für alle* ($\tilde{\mu}^{(0)}$-) *meßbaren* $X \subseteqq X^{(0)}$ $\tilde{\mu}^{(0)}(X) = \mu\left(\sum_{x \text{ in } X} M(x)\right)$. *Es ist stets* $\mu_{(x)}(M(x)) = 1$.

Es ist für alle (μ-) *meßbaren* $N \subseteq \Omega_1$ $\mu(N) = \int_X \mu_{(x)}(M(x) \cdot N)\, dw(x)$.

(*D. h.: der Integrand ist, bis auf eine x-Menge vom* ($\tilde{\mu}^{(0)}$-) *Maße* 0, *sinnvoll, also* $M(x) \cdot N$ ($\mu_{(x)}$-) *meßbar;* $\mu_{(x)}(M(x) \cdot N)$ *ist als x-Funktion* ($\tilde{\mu}^{(0)}$-) *meßbar;* $dw(x)$ *ist das Differentialelement der* $\tilde{\mu}^{(0)}$*-Integration, also*

$$w(y) = \tilde{\mu}^{(0)}(\text{Menge } 0 \leqq x \leqq y) = \mu\left(\sum_{0 \leqq x \leqq y} M(x)\right);$$

und außerdem wird die Gleichheit beider Seiten behauptet.)

Damit haben wir bewiesen, was in Einleitung 4. angekündigt wurde: daß jede allgemeine Strömung in ergodische Strömungen zerlegt werden kann. Es ließe sich noch zeigen, daß die Zerlegung von Satz 2 „im Wesentlichen" nur auf eine Weise möglich ist, wir gehen aber darauf nicht mehr ein.

III. Charakteristische Eigenschaften der Koopmanschen U_t-Scharen.

1. Betrachten wir zunächst eine einzige eineindeutige maßtreue Abbildung S: $P \to SP$ von Ω' auf Ω'', oder auch nur von Ω' minus einer Menge vom Maße 0 auf Ω'' minus eine Menge vom Maße 0. Daß der durch $Uf(P) = f(SP)$ definierte Operator U unitär ist, ist klar, ebenso daß (D_s entsprechend) $Uf \cdot Ug = U(f \cdot g)$ ist, falls $f, g, f \cdot g$ zum Hilbertschen Raume $\mathfrak{H}$ gehören. Letzteres ist z. B. bestimmt der Fall, wenn f, g zum Hilbertschen Raume gehören und beschränkt sind[55]. Hat nun umgekehrt ein unitärer Operator U in $\mathfrak{H}$ diese Eigenschaft: $Uf \cdot Ug = U(f \cdot g)$ für beschränkte f, g, so können wir ein eineindeutiges und maßtreues S finden, das U nach $Uf(P) = f(SP)$ erzeugt.

Sei nämlich $A' \subseteqq \Omega$ meßbar und von endlichem Maße, dann ist für $\chi_{A'}(P) \begin{cases} = 1 \text{ für } P \text{ in } A' \\ = 0 \quad \text{sonst} \end{cases}$ $\chi_{A'} \cdot \chi_{A'} = \chi_{A'}$, also auch $U\chi_{A'} \cdot U\chi_{A'} = U(\chi_{A'} \cdot \chi_{A'}) = U\chi_{A'}$. Somit ist, bis auf eine Menge vom Maße 0, $U\chi_{A'}(P) = 0$ oder 1, und da wir es auf dieser etwa zu 0 umdefinieren können, ohne es als $\mathfrak{H}$-Element zu ändern, dürfen wir annehmen, daß $U\chi_{A'}(P) = 0$ oder 1 stets gilt. Die Menge, wo letzteres gilt, heiße A'', sie ist meßbar (weil $U\chi_{A'}$ es ist) und von endlichem Maße (wie $U\chi_{A'}$ zu $\mathfrak{H}$ gehört). Es ist somit $U\chi_{A'} = \chi_{A''}$.

[55] Denn es ist

$$\int_\Omega |f(P)\, g(P)|^2 \, dv_P \leqq \left(\operatorname*{Max}_{P \text{ in } \Omega} |f(P)|\right)^2 \int_\Omega |g(P)|^2 \, dv_P,$$

also endlich.

Wir haben durch diese Zuordnung jedem meßbaren $\Lambda' \subseteq \Omega'$ von endlichem Maße ein ebensolches $\Lambda'' \subseteq \Omega''$ zugeordnet; da U^{-1} ebenso ist, gilt auch die Umkehrung, d. h. jedes solche Λ'' tritt bei einem gewissen Λ' wirklich auf. Ferner legt Λ' offenbar Λ'' bis auf eine Menge vom Maße 0 fest, ebenso Λ'' das Λ'.

Da U unitär ist und $\|\chi_{\Lambda'}\|^2 = \mu(\Lambda')$, $\|\chi_{\Lambda''}\|^2 = \mu(\Lambda'')$, ist $\mu(\Lambda') = \mu(\Lambda'')$. Es ist

$$U\chi_{\Lambda_1'\cdot\Lambda_2'} = U(\chi_{\Lambda_1'}\cdot\chi_{\Lambda_2'}) = U\chi_{\Lambda_1'}\cdot U\chi_{\Lambda_2'} = \chi_{\Lambda_1''}\cdot\chi_{\Lambda_2''} = \chi_{\Lambda_1''\cdot\Lambda_2''}$$

und

$$U\chi_{\Lambda_1'+\Lambda_2'} = U(\chi_{\Lambda_1'}+\chi_{\Lambda_2'}-\chi_{\Lambda_1'\cdot\Lambda_2'}) = U\chi_{\Lambda_1'}+U\chi_{\Lambda_2'}-U\chi_{\Lambda_1'\cdot\Lambda_2'}$$
$$= \chi_{\Lambda_1''}+\chi_{\Lambda_2''}-\chi_{\Lambda_1''\cdot\Lambda_2''} = \chi_{\Lambda_1''+\Lambda_2''}.$$

D. h.: wenn Λ_1', Λ_2' bzw. Λ_1'', Λ_2'' entsprechen, entspricht $\Lambda_1'\cdot\Lambda_2'$ $\Lambda_1''\cdot\Lambda_2''$ und $\Lambda_1'+\Lambda_2'$ $\Lambda_1''+\Lambda_2''$. Durch Induktion folgt: wenn $\Lambda_1', \cdots, \Lambda_n'$ bzw. $\Lambda_1'', \cdots, \Lambda_n''$ entsprechen, so entspricht $\Lambda_1'+\cdots+\Lambda_n'$ bzw. $\Lambda_1''+\cdots+\Lambda_n''$. Und da U ein stetiger Operator ist, ergibt ein Grenzübergang: wenn $\Lambda_1', \Lambda_2', \cdots$ bzw. $\Lambda_1'', \Lambda_2'', \cdots$ entsprechen (und $\Lambda_1'+\Lambda_2'+\cdots$ ein endliches Maß hat), so entspricht $\Lambda_1'+\Lambda_2'+\cdots$, $\Lambda_1''+\Lambda_2''+\cdots$.

Wir haben somit in $\Lambda' \leftrightarrow \Lambda''$, im Sinne von Satz 1 der vorhergehenden Abhandlung des Verfassers, eine maßtreue Mengenabbildung von Ω' auf Ω'' vor uns, und ebenso ist $\Lambda'' \leftrightarrow \Lambda'$. Dann existiert aber auf Grund des genannten Satzes eine eineindeutige maßtreue Punktabbildung $P \to SP$ von Ω' minus einer Menge vom Maße 0 auf Ω'' minus eine Menge vom Maße 0, die diese Mengenabbildung $\Lambda'' \leftrightarrow \Lambda'$ erzeugt. Wenn Ω', Ω'' unabzählbar sind, können wir S sogar Ω' eineindeutig auf Ω'' abbildend wählen.

$Vf(P) = f(SP)$ ist also unitär. Es ist $U\chi_{\Lambda'} = \chi_{\Lambda''}$, aber Λ'' ist das $P \to SP$-Urbild von Λ', also auch $V\chi_{\Lambda'} = \chi_{\Lambda''}$. Somit gilt $Uf = Vf$ für alle $f = \chi_{\Lambda'}$. Da aber diese die abgeschlossene Linearmannigfaltigkeit $\mathfrak{H}$ aufspannen (vgl. a. a. O. Anm. [45]), gilt es für alle f. Also ist $U = V$, womit alles bewiesen ist.

2. Sei nun S_t eine allgemeine Strömung, U_t die Koopmanschen Operatoren: $U_t f(P) = f(S_t P)$. Unitarität, $U_t U_s = U_{t+s}$, sowie $U_t f \cdot U_t g = U_t(f \cdot g)(f, g$ beschränkt) sind klar, es fehlt nur noch eins aus D_4, D_5, nämlich daß $(U_t f, g)$ eine meßbare Funktion von t ist. Dies genügt es aber, für die f, g aus einer solchen Menge $\mathfrak{K}$ zu beweisen, die die abgeschlossene Linearmannigfaltigkeit $\mathfrak{H}$ aufspannt, denn wenn $(U_t f, g)$ für die f, g von $\mathfrak{K}$ meßbar ist, ist es es auch für deren Linearaggregate und die Limites dieser Linearaggregate, d. h. für alle f, g. Eine solche Menge $\mathfrak{K}$ ist nun die Menge aller χ_Λ (Λ meßbar, von endlichem Maß, vgl. a. a. O. Anm. [45]), wir haben also nur zu zeigen: $(U_t\chi_\Lambda, \chi_M) = (\chi_{S_t^{-1}\Lambda}, \chi_M) = \mu(S_t^{-1}\Lambda \cdot M)$ ist eine

meßbare Funktion von t. In II. 6. zeigten wir, daß das $[P, t] \to S_t P$-Urbild von $\mathcal{A}$ (in $\Omega \times T$) meßbar ist, und die Menge der $[P, t]$, P in M, ist offenbar auch meßbar. Folglich ist der Durchschnitt dieser zwei Mengen, $\bar{N}$, meßbar. Nach dem Satz von Fubini ist also $\mu(\bar{N}_{(t)})$ ($\bar{N}_{(t)}$ ist die P-Menge mit $[P, t]$ in $\bar{N}$) eine meßbare Funktion von t. Da aber $\bar{N}_{(t)} = S_t^{-1}\mathcal{A} \cdot M$ ist, sind wir damit am Ziel.

Schlagen wir nun den umgekehrten Weg ein: sei U_t eine D_4, D_5 erfüllende Operatorenschar. Zunächst können wir nach 1. jedem U_t ein $P \to S_t P$ zuordnen, das Ω minus eine Menge vom Maße 0 eineindeutig und maßtreu auf Ω minus eine Menge vom Maße 0 abbildet. Aus $U_t U_s = U_{t+s}$ folgt $S_t S_s \cong S_{t+s}$. Es bleibt noch zu prüfen, ob $S_t P$ meßbar von $[P, t]$ abhängt.

Aus Satz 1 der vorhergehenden Abhandlung des Verfassers folgt, daß U_t stark stetig in t ist, also insbesondere $U_t f \to f$ für $t \to 0$. Ist $f = \chi_\Delta$, so besagt dies $\chi_{S_t^{-1}\Delta} \to \chi_\Delta$, und da $\| \chi_{S_t^{-1}\Delta} - \chi_\Delta \|^2 = \mu((\mathcal{A} + S_t^{-1}\mathcal{A}) - (\mathcal{A} \cdot S_t^{-1}\mathcal{A}))$ ist, $\mu((\mathcal{A} + S_t^{-1}\mathcal{A}) - (\mathcal{A} \cdot S_t^{-1}\mathcal{A})) \to 0$. Wenn von P, $S_t P$ beide oder keins in $\mathcal{A}$ liegen, sagen wir, sie liegen in bezug auf $\mathcal{A}$ gleich; wie wir sehen, strebt das Maß der Menge der mit ihrem S_t-Bilde in bezug auf $\mathcal{A}$ nicht gleich liegenden P mit $t \to 0$ gegen Null. Sei nun $K_1, K_2, \cdots$ ein „gleichwertiges Umgebungssystem" in Ω. Sei $\varepsilon > 0$ und ein $n = 1, 2, \cdots$ gegeben. Das Maß der Menge der mit ihrem S_t-Bilde in bezug auf K_m nicht gleich liegenden P strebt mit $t \to 0$ gegen Null, also auch das Maß der Summe dieser Mengen für alle $m = 1, \cdots, n$. Es gibt also ein $\delta = \delta(n, \varepsilon) > 0$, so daß dieses Maß $\leqq \varepsilon$ ist, wenn $|t| \leqq \delta$ ist. Sei nun $|t'' - t'| \leqq \delta$. Dann hat die Menge der P, die mit ihrem $S_{t''-t'}$-Bilde nicht in bezug auf alle $K_1, \cdots, K_n$ gleich liegen, ein Maß $\leqq \varepsilon$, also auch ihr $S_{t'}$-Urbild, d. i. die Menge $\Delta(t', t'', n)$ aller P, für welche das $S_{t'}$- und das $S_{t''}$-Bild nicht in bezug auf alle $K_1, \cdots, K_n$ miteinander gleich liegen.

Sei $\delta_n = \delta\left(n, \frac{1}{2^n}\right)$ und eine Folge $t_1, t_2, \cdots$ sowie ein t gegeben mit $|t - t_n| \leqq \delta_n$. Die Menge $\Delta_n = \sum_{m=n}^{\infty} (t_m, t, m)$ hat ein Maß $\leqq \sum_{m=n}^{\infty} \frac{1}{2^m} = \frac{1}{2^{n-1}}$. Liegt P nicht in ihr, so liegt $S_{t_n} P$ mit $S_t P$ in bezug auf alle $K_1, \cdots, K_n$ gleich. Da dann P erst recht in keinem Δ_p, $p \geqq n$ liegt, liegen die $S_{t_p} P$ mit $S_t P$ in bezug auf alle $K_1, \cdots, K_p$ gleich, also erst recht in bezug auf alle $K_1, \cdots, K_n$. D. h.: alle $S_{t_p} P$, $p \geqq n$ sowie $S_t P$ liegen miteinander in bezug auf alle $K_1, \cdots, K_n$ gleich.

Liege nun P nicht in Δ_n, sei K eine Umgebung von $S_t P$. Dann gibt es ein $S_t P$ enthaltendes $K_m \subseteq K$. Sei $p = \mathrm{Max}(m, n)$, dann liegen nach Obigem alle $S_{t_q} P$, $q \geqq p$, mit $S_t P$ in bezug auf alle $K_1, \cdots, K_p$ gleich (denn P gehört erst recht nicht zu Δ_p), also auch in bezug auf K_m. Da $S_t P$ in K_m liegt, liegen somit auch die $S_{t_q} P$ dort, also erst recht in K. D. h.: alle

$S_{t_q}P$ mit $q \geqq p$ liegen in K. Da dies für jedes K gilt, haben wir $S_{t_q}P \to S_tP$ für $q \to \infty$. Die Menge der P, für die nicht $S_{t_q}P \to S_tP$ für $q \to \infty$ ist, ist also jedenfalls $\subseteqq \Delta_n$. Ihr (äußeres) Maß ist somit $\leqq \frac{1}{2^{n-1}}$. Und da dies für jedes n gilt, hat sie das Maß 0.

Wir haben gezeigt: wenn $|t - t_n| \leqq \delta_n$ ist, so gilt, bis auf eine P-Menge vom Maße 0, $S_{t_n}P \to S_tP$ für $n \to \infty$. Nun definieren wir für jedes $n = 1, 2, \cdots$: $\Sigma_t^{(n)}P = S_{\nu\delta_n}P$, wobei $\nu = \nu(n, t)$ aus $\nu\delta_n \leqq t < (\nu+1)\delta_n$ zu bestimmen ist. In einem Intervalle $\nu_0\delta_n \leqq t < (\nu_0+1)\delta_n$ ist $\Sigma_t^{(n)}P$ von t unabhängig, und in seiner P-Abhängigkeit ein $S_{t_0}P(t_0 = \nu_0\delta)$, also sowohl P- als auch $[P, t]$-meßbar. Da aber eine Folge solcher Intervalle den ganzen $[P, t]$-Raum $(\Omega \times T)$ ausmacht, ist es überhaupt $[P, t]$-meßbar. Wir können nun ein $S_t'P$ so definieren: wenn $\lim\limits_{n\to\infty} \Sigma_t^{(n)}P$ existiert, sei $S_t'P$ dieser Limes, sonst sei es undefiniert. Dieses $S_t'P$ ist offenbar auch meßbar. Da nun $|t - \nu\delta_n| \leqq \delta_n$ ist, folgt aus dem weiter oben Bewiesenen (mit $t_n = \nu(n,t)\delta_n$), daß bei festem t $S_t'P$ bis auf eine P-Menge vom Maße 0 definiert und gleich S_tP ist. Somit bilden auch die S_t' Ω minus eine Nullmenge eineindeutig und maßtreu auf Ω minus eine Nullmenge ab, auch sie erzeugen die unitären Operatoren U_t, auch für sie gilt $S_t'S_s' = S_{t+s}'$. Damit D_2' erfüllt sei, ist nur noch eines notwendig: $S_t'P$ in jenen P-Mengen vom Maße 0 undefiniert machen, die bei der eineindeutigen Abbildung $P \to S_tP$ zu streichen sind — und dies unter Wahrung seiner $[P, t]$-Meßbarkeit.

Daß das $[P, t] \to [S_t'P, t]$-Urbild jeder meßbaren Menge $\bar{N}$ meßbar ist, zeigt man wie in 6., also ist es auch das $[P, t] \to S_t'P$-Urbild jeder meßbaren Menge N (man setze $\bar{N} = N \times T$). Daher ist bei festem s $S_sS_t'P$ eine meßbare $[P, t]$-Funktion. Da $\Sigma_{-t}^{(n)}S_t'P$ aus abzählbar vielen solchen zusammengesetzt ist, ist es ebenfalls $[P, t]$-meßbar, und sein Limes für $n \to \infty$, $S_{-t}'S_t'P$ ist es auch. Die $[P, t]$-Menge $\bar{M}$ mit $S_{-t}'S_t'P \neq P$ ist also meßbar, und für jedes t hat die Menge der P mit $[P, t]$ in $\bar{M}$ das Maß 0. Nach dem Satze von Fubini hat also auch $\bar{M}$ das Maß 0. Erklären wir daher $S_t'P$ in $\bar{M}$ für undefiniert, so sind wir am Ziele.

Wir haben bewiesen:

SATZ 3. *Damit eine Operatorenschar nach Einleitung* 3. *zu einer allgemeinen Strömung S_t gehöre, d. h. damit sie eine Koopmansche Schar sei, ist notwendig und hinreichend, daß sie D_4, D_5 (Einleitung* 3.) *erfüllt.* (D_5 *braucht bloß für beschränkte f, g gefordert zu werden.*)

3. Wir holen jetzt noch den Beweis der in II., 8. verwendeten Aussage nach. Es sei eine Schar eineindeutiger und maßtreuer Abbildungen S_t von Ω minus einer Menge vom Maße 0 auf Ω minus eine Menge vom Maße 0 gegeben, jedoch sei S_t nur für die t außerhalb einer gewissen t-Menge T'

vom (gewöhnlichen Lebesgueschen) Maße 0 definiert, und $S_t S_s \cong S_{t+s}$ braucht auch bei festem t nur bis auf eine s-Menge vom Maße 0 zu gelten, schließlich sei $S_t P$ als $[P, t]$-Funktion meßbar. Behauptet wird: es ist möglich, die S_t für die t von T' so hinzu zu definieren, daß eine allgemeine Strömung entsteht (d. h. D_2' erfüllt ist).

Zu den gegebenen S_t (t nicht in T') können wir jedenfalls die U_t bilden, sie sind unitär, und für jedes gegebene t (nicht in T') gilt für alle s, bis auf eine s-Menge vom Maße 0, $U_t U_s = U_{t+s}$. Ferner ist $(U_t f, g)$ meßbar in t, was genau so bewiesen wird, wie am Anfang von 2. Könnten wir nun für die t von T' die U_t so hinzudefinieren, daß ausnahmslos $U_t U_s = U_{t+s}$ gilt, so erfüllen die neuen U_t D_4, D_5, denn auch die t-Meßbarkeit von $(U_t f, g)$ bleibt erhalten, da die Neudefinition in einer t-Menge vom Maße 0 erfolgte. Nach Satz 3 gibt es dann eine allgemeine Strömung S_t', die diese U_t erzeugt. Für die t außerhalb von T' erzeugen S_t, S_t' dasselbe U_t, es ist also $S_t \cong S_t'$. Wenn wir also S_t' außerhalb von T' in S_t umdefinieren, entsteht eine äquivalente allgemeine Strömung, S_t'', und diese ist die gesuchte Erweiterung von S_t.

Die Aufgabe ist also: T' hat das Maß 0; für jedes t außerhalb T' ist ein unitäres U_t gegeben, und es ist für alle s, bis auf eine (von t abhängige) s-Menge vom Maße 0, $U_t U_s = U_{t+s}$; $(U_t f, g)$ ist eine meßbare t-Funktion — für die t von T' soll U_t so hinzudefiniert werden, daß ausnahmslos $U_t U_s = U_{t+s}$ gilt.

Da stets $1 \cdot U_s = U_{0+s}$ ist, ist U_0, falls 0 nicht zu T' gehört, gleich 1; und wenn 0 zu T' gehört, können wir es aus T' herausnehmen und $U_0 = 1$ definieren. 0 nicht in T', $U_0 = 1$, darf also jedenfalls vorausgesetzt werden.

Bei gegebenem t bilden die r von T', sowie die r mit $t+r$ in T' je eine Menge vom Maße 0, wir können also r so wählen, daß weder r, noch $s = t+r$ zu T' gehört. D. h.: s, r nicht in T', $t = s-r$, kann erreicht werden. Wir setzen nun $U_t' = U_s U_r^{-1}$, wobei freilich noch zu zeigen ist, daß U_t' von der Wahl von s, r (innerhalb der obigen Bedingungen) nicht abhängt. Ist aber dies gezeigt, so schließen wir so weiter: Für t nicht in T' können wir r, $t+r$ außerhalb von T' wählen, so daß $U_t U_r = U_{t+r}$ gilt, dann ist $U_t = U_{t+r} U_r^{-1} = U_t'$. Für beliebige t, s können wir r, $s+r$, $t+s+r$ außerhalb von T' wählen, dann ist

$$U_t' U_s' = U_{t+s+r} U_{s+r}^{-1} \cdot U_{s+r} U_r^{-1} = U_{t+s+r} U_r^{-1} = U_{t+s}'.$$

D. h. U_t' ist die gesuchte Definitionserweiterung für U_t. Es ist also nur noch zu zeigen: aus s, r, s', r' nicht in T', $s-r = s'-r'$ folgt $U_s U_r^{-1} = U_{s'} U_{r'}^{-1}$.

Es genügt zu zeigen: aus s, r, s', r' nicht in T', $s+r=s'+r'$ folgt $U_s U_r = U_{s'} U_{r'}$. Denn mit $r'=s$, $s'=r$ ergibt das $U_s U_r = U_r U_s$, d. h. die U_t kommutieren miteinander. Und damit wird $U_s U_r = U_{s'} U_{r'}$ $U_s U_{r'}^{-1} = U_{s'} U_r^{-1}$ gleichbedeutend. $s+r=s'+r'$ ist $s-r'=s'-r$ gleichbedeutend, wir haben also die ursprüngliche Behauptung wieder, bloß mit s, r', s', r an Stelle von s, r, s', r'.

Um die obige Behauptung zu beweisen, betrachten wir den Beweis von Satz 1 in der zweitvorhergehenden Abhandlung des Verfassers. Der Beweis dafür, daß für alle t außerhalb einer Menge T'' ($\supseteq T'$) vom Maße 0 U_t approximativ stark-stetig ist (vgl. a. a. O.), gilt unverändert. Gehört t nicht zu T', so ist für alle s, bis auf eine Menge vom Maße 0, $U_t U_s = U_{t+s}$, also

$$\|U_{t+s} f - U_t f\| = \|U_t U_s f - U_t U_0 f\| = \|U_t (U_s f - U_0 f)\| = \|U_s f - U_0 f\|.$$

Wenden wir dies auf ein t außerhalb von T'' an, so ergibt es, daß U_t insbesondere für $t=0$ approximativ stark-stetig ist; wenden wir es auf die t von T' an, so zeigt sich, daß aus der approximativen starken Stetigkeit von U_t für $t=0$ dieselbe für alle t von T' folgt — d. h. $T'=T''$.

Seien nun s, r, s', r' gegeben, nicht in T', $s+r=s'+r'$. Sei f aus $\mathfrak{H}$, $\varepsilon>0$. Dann gibt es (da U_t für $t=r$ und $t=r'$ approximativ stark-stetig ist) ein $\delta_1>0$, so daß für jedes $\delta \leqq \delta_1$, >0 in $|t-r| \leqq \delta$ $\|U_t f - U_r f\| \leqq \frac{\varepsilon}{2}$ gilt, bis auf eine Menge vom Maße $<\delta$, ebenso ein $\delta_2>0$, so daß für jedes $\delta \leqq \delta_2$, >0 in $|t'-r'| \leqq \delta$ $\|U_{t'} f - U_{r'} f\| \leqq \frac{\varepsilon}{2}$ gilt, bis auf eine Menge vom Maße $<\delta$. Ferner gilt bis auf Mengen vom Maße 0 $U_s U_t = U_{s+t}$, sowie $U_{s'} U_{t'} = U_{s'+t'}$. Sei $\delta = \mathrm{Min}(\delta_1, \delta_2)$. Binden wir also t, t' durch die Bedingung $s+t=s'+t'$ aneinander, so gelten, bis auf eine Menge vom Maße $<2\delta$ alle vier Bedingungen. In $|t-r|<\delta$, das $|t'-r'|<\delta$ gleichbedeutend ist, und das Maß 2δ hat, gibt es also ein t, das sie alle erfüllt. Also ist

$$U_s U_t = U_{s+t} = U_{s'+t'} = U_{s'} U_{t'},$$

$$\|U_s U_t f - U_s U_r f\| = \|U_s (U_t f - U_r f)\| = \|U_t f - U_r f\| \leqq \frac{\varepsilon}{2},$$

$$\|U_{s'} U_{t'} f - U_{s'} U_{r'} f\| = \|U_{s'} (U_{t'} f - U_{r'} f)\| = \|U_{t'} f - U_{r'} f\| \leqq \frac{\varepsilon}{2},$$

also $\|U_s U_r f - U_{s'} U_{r'} f\| \leqq \varepsilon$. Da dies für jedes $\varepsilon>0$ gilt, ist $U_s U_r f = U_{s'} U_{r'} f$, und da dies für jedes f gilt, $U_s U_r = U_{s'} U_{r'}$, wie behauptet wurde.

Damit ist der notwendige Beweis nachgeholt.

IV. Die U_t-Scharen mit Punktspektrum.

1. In Einleitung 3 haben wir den auf dem Satz von Stone beruhenden Zusammenhang zwischen der U_t-Schar, der Zerlegung der Einheit $E(\lambda)$, und dem hypermaximalen Operator A angegeben. Es ist bekannt, was unter dem Spektrum von A zu verstehen ist, und wie dieses eingeteilt wird[57], wir wollen es auch Spektrum der $E(\lambda)$ oder Spektrum der U_t nennen, da es diese auch festlegen. Im Folgenden sollen die Eigenschaften dieses Spektrums untersucht werden.

Auf Grund von Satz 2 in § II brauchen wir nur solche U_t-Scharen zu untersuchen, die zu ergodischen Strömungen gehören. D. h. für die bei $\lim\limits_{\substack{\varepsilon>0\\ \varepsilon\to 0}} (E(+\varepsilon)-E(-\varepsilon)) = \lim\limits_{\substack{\varepsilon>0\\ \varepsilon\to 0}} (E(0)-E(-\varepsilon))$ ($E(\lambda)$ ist nach rechts halbstetig!) $= E_0$ nur für konstantes f $E_0 f = f$ gilt. Bei Null-ergodischer Strömung ($\mu(\Omega) = \infty$) gehört außer 0 kein konstantes f zu $\mathfrak{H}$, also ist $E_0 = 0$; bei eigentlich-ergodischer Strömung ($\mu(\Omega)$ endlich) gehören die Konstanten f zu $\mathfrak{H}$, sie sind die $c\varphi_0$, wo φ_0 das normierte $\frac{1}{\sqrt{\mu(\Omega)}}$ ist, also ist $E_0 = P_{[\varphi_0]}$[58]. Im ersten Falle gehört also 0 nicht zum Punktspektrum; im zweiten gehört es dazu, u. zw. als einfacher Eigenwert mit der Eigenfunktion $\varphi_0 = \frac{1}{\sqrt{\mu(\Omega)}}$.

Der Null-ergodische Fall wurde von Koopman diskutiert (a. a. O. Anm. [2]). Gäbe es dort ein Punktspektrum, so wäre für ein $f \neq 0$ $Af = \lambda f$, d. h. $U_t f = e^{i\lambda t} f$. Hieraus folgt

$$U_t |f(P)|^{59} = |U_t f(P)| = |e^{it\lambda} \cdot f(P)| = |f(P)|.$$

Also ist $|f(P)|$ Eigenfunktion vom Eigenwerte 0, da aber 0 kein Eigenwert ist, $|f(P)| = 0$, $f(P) = 0$. Es gibt also kein Punktspektrum, sondern nur ein reines Streckenspektrum.

Im eigentlich-ergodischen Falle können wir $\mu(\Omega) = 1$ annehmen (vgl. Satz 2). Wieder folgt nach Koopman für jede Eigenfunktion f des Eigenwertes λ, daß $|f|$ Eigenfunktion des Eigenwertes 0 ist. Hier bedeutet dies aber nur, das $|f|$ konstant ist, also $|f| = c > 0$, $f(P) = ce^{i\theta(P)}$. Und $U_t f(P) = f(S_t P) = e^{i\lambda t} f(P)$ bedeutet $\theta(S_t P) = \theta(P) + \lambda t \cdots \bmod 2\pi$. $\theta(P)$ ist offenbar überhaupt nur $\bmod 2\pi$ definiert, reell, und (wenn man es z. B. auf Werte $\geqq 0$, $< 2\pi$ reduziert) meßbar. Hat umgekehrt ein $\theta(P)$ diese

[57] Hilbert, Gött. Nachr. 1906, S. 157; Hellinger, Dissertation, Göttingen, 1907.

[58] Vgl. die Theorie der Projektionsoperatoren in E, S. 74—78.

[59] Es ist offenbar allgemein $U_t F(f(P)) = F(U_t f(P))$. Mit $f(P)$ gehört auch $|f(P)|$ zu $\mathfrak{H}$.

drei Eigenschaften, so gehört jedes $f(P) = ce^{i\theta(P)}$ zu $\mathfrak{H}$, und erfüllt $U_t f(P) = f(S_t P) = e^{i\lambda t} f(P)$, d. h. es ist Eigenfunktion vom Eigenwerte λ. Es gilt also:

SATZ 4. (Koopman, a. a. O. Anm. [2]) *Sei Ω eigentlich-ergodisch. $f(P)$ ist dann und nur dann Eigenfunktion vom Eigenwert λ, wenn es die Form $ce^{i\theta(P)}$ hat. Hier ist $\theta(P)$ eine* mod 2π *definierte, reelle, meßbare Funktion, für die $\theta(S_t P) = \theta(P) + \lambda t$ gilt.*

Zwei Eigenfunktionen f gelten als nicht wesentlich verschieden, wenn sie sich nur um einen konstanten Faktor unterscheiden; also zwei θ, wenn sie sich nur um einen konstanten Addenden unterscheiden. Für $\lambda = 0$ muß, wegen des ergodischen Charakters, $f =$ Konstans sein, also auch $\theta =$ Konstans, d. h. im wesentlichen $\theta = 0$.

Wir wollen die $\theta(P)$ von Satz 4, mit Hinblick auf die entsprechende Begriffsbildung der klassischen Mechanik, Winkelvariable nennen, u. zw. zum Eigenwert λ gehörige.

Einige ihrer Eigenschaften liegen auf der Hand: Sind θ_1, θ_2 Winkelvariable vom Eigenwert λ, so ist $\theta_1 - \theta_2$ eine vom Eigenwert 0, d. h. konstant, also $\theta_1 = \theta_2 +$ Konstans, d. h.: zu einem gegebenen λ gehört keine oder (im wesentlichen) genau eine Winkelvariable. Sind θ_1, θ_2 Winkelvariable der bzw. Eigenwerte λ_1, λ_2, so ist $\theta_1 + \theta_2$ Winkelvariable des Eigenwertes $\lambda_1 + \lambda_2$. Die Winkelvariablen und ihre Eigenwerte bilden also einen „Additiv-Modul", die letzteren sind aber das Punktspektrum der U_t: dieses enthält somit mit λ_1, λ_2 stets auch $\lambda_1 \pm \lambda_2$. Und nach dem vorher Gesagten hat die U_t-Schar nur einfache Punkteigenwerte.

Sind θ_1, θ_2 (wesentlich) verschiedene Winkelvariable (also mit verschiedenen Eigenwerten), so sind die Eigenfunktionen $e^{i\theta_1}$, $e^{i\theta_2}$ orthogonal, also

$$\int_\Omega e^{i\theta_1(P)} e^{\overline{i\theta_2(P)}}\, d\,v_P = \int_\Omega e^{i(\theta_1(P) - \theta_2(P))}\, d\,v_P = 0.$$

2. Gehen wir nun umgekehrt vor. Sei ein abzählbar unendliches Zahlensystem Z gegeben, das mit λ_1, λ_2 auch $\lambda_1 \pm \lambda_2$ enthält. Gibt es eine eigentlich-ergodische U_t-Schar mit reinem Punktspektrum (das, wie oben gezeigt wurde, einfach sein muß), so daß dieses Punktspektrum gleich Z ist?

Um eine solche U_t-Schar anzugeben, genügt es, ein vollständiges normiertes Orthogonalsystem beschränkter φ_λ (λ durchläuft Z) zu finden, das die Eigenschaft $\varphi_\lambda \cdot \varphi_\mu = \varphi_{\lambda+\mu}$ besitzt. Denn da jedes f als $\sum_{\lambda \text{ in } Z} a_\lambda \varphi_\lambda$ geschrieben werden kann, und dann $\|f\|^2 = \sum_{\lambda \text{ in } Z} |a_\lambda|^2$ ist (vollständiges normiertes

Orthogonalsystem!), können wir Operatoren U_t durch $U_t\,(\sum_{\lambda \text{ in } Z} a_\lambda\,\varphi_\lambda) = \sum_{\lambda \text{ in } Z} e^{i\lambda t}\,a_\lambda\,\varphi_\lambda$ definieren, dieselben sind unitär. $U_t\,U_s = U_{t+s}$ ist klar. Ebenso $U_t f \cdot U_t g = U_t f \cdot g$ für $f = \varphi_\lambda$, $g = \varphi_\mu$, daher gilt es auch für alle f, die Linearaggregate von φ_λ-s sind, und deren Limites (die Beschränktheit von $g = \varphi_\mu$ erlaubt den Stetigkeitsschluß!), d. h. alle f. Ist f beschränkt, so überträgt es sich ebenso auf die Linearaggregate der φ_μ, und deren Limites (Stetigkeitsschluß!, vgl. vorhin). Somit gelten D_4, D_5, und nach Satz 3 entsteht die Schar U_t wirklich aus einer allgemeinen Strömung S_t. Ferner sieht man, daß $U_t\,\varphi_\lambda = e^{it\lambda}\,\varphi_\lambda$ ist, also ist φ_λ Eigenfunktion vom Eigenwert λ. Da die φ_λ ein vollständiges normiertes Orthogonalsystem bilden, gibt es keine weiteren Eigenfunktionen. Somit hat die U_t-Schar ein reines Punktspektrum, das mit Z zusammenfällt und einfach ist. Da $0 = \lambda - \lambda$ bestimmt zu Z gehört, ist 0 einfacher Eigenwert. Die Eigenfunktion von 0, φ_0, ist wegen $\varphi_0 \cdot \varphi_\mu = \varphi_\mu$ gleich 1, d. h. konstant. Somit ist S_t eigentlich-ergodisch, und wir sind am Ziele.

Daß aber das vollständige normierte Orthogonalsystem φ_λ, λ in Z, mit den geforderten Eigenschaften existiert, ist die Aussage eines schönen Satzes von A. Haar[60].

3. Seien nun zwei m-Räume Ω', Ω'' mit bzw. l-Maßen μ'^*, μ''^* gegeben, und den bzw. allgemeinen Strömungen S'_t, S''_t. Diese Strömungen seien eigentlich-ergodisch, $\mu(\Omega') = \mu(\Omega'') = 1$. Die entsprechenden Operatorenscharen U'_t, U''_t mögen reine Punktspektra besitzen, und zwar beide dasselbe: Z.

Da beide Spektra einfach sein müssen, gehört zu jedem λ von Z genau je eine normierte Eigenfunktion dieses Eigenwertes bei U'_t bzw. U''_t: φ'_λ bzw. φ''_λ (diese sind nur bis auf konstante Faktoren vom Absolutwert 1 festgelegt, seien aber jetzt fest gewählt). Sowohl die φ'_λ als auch die φ''_λ bilden je ein vollständiges normiertes Orthogonalsystem.

φ'_λ hat einen konstanten Absolutwert c'_λ, also ist $\|\varphi'_\lambda\|^2 = c'^2_\lambda \cdot \mu(\Omega)$, und da die linke Seite, sowie $\mu(\Omega)$ gleich 1 ist, ist $c'_\lambda = 1$, d. h. $|\varphi'_\lambda| = 1$. Also ist auch $|\varphi'_{\lambda_1} \cdot \varphi'_{\lambda_2}| = 1$. Da aber $\varphi'_{\lambda_1} \cdot \varphi'_{\lambda_2}$ Eigenfunktion vom Eigenwert $\lambda_1 + \lambda_2$ ist, muß $\varphi'_{\lambda_1} \cdot \varphi'_{\lambda_2} = c'_{\lambda_1,\lambda_2} \cdot \varphi'_{\lambda_1+\lambda_2}$, c'_{λ_1,λ_2} konstant, gelten; und da auch $|\varphi'_{\lambda_1+\lambda_2}| = 1$ ist, muß $|c'_{\lambda_1,\lambda_2}| = 1$ sein. Ebenso zeigt man $\varphi''_{\lambda_1} \cdot \varphi''_{\lambda_2} = c''_{\lambda_1,\lambda_2} \cdot \varphi''_{\lambda_1+\lambda_2}$, $c''_{\lambda_1,\lambda_2}$ konstant, $|c''_{\lambda_1,\lambda_2}| = 1$.

Wir wollen nun die φ'_λ durch Multiplikation mit geeigneten Konstanten vom Absolutwerte 1 in ein neues vollständiges normiertes Orthogonal-

[60] Math. Zeitschr. 33, 1931, S. 132, Satz II, und zwar ist Z der dortigen kommutativen Gruppe $A, B, C, \ldots$ gleichzusetzen, $+$ ist das Kompositionsgesetz der Gruppe, unsere $\varphi_\lambda(P)$, λ in Z, entsprechen den dortigen $X_A(s)$, $X_B(s)$, ...

system von U_t-Eigenfunktionen $\psi'_\lambda = d'_\lambda \varphi'_\lambda$ überführen, so daß $k_1 \lambda_1 + \cdots + k_n \lambda_n = 0$ ($\lambda_1, \cdots, \lambda_n$ aus Z, $k_1, \cdots, k_n = 0, \pm 1, \pm 2, \cdots$) $(\psi'_{\lambda_1})^{k_1} \cdot \ldots \cdot (\psi'_{\lambda_n})^{k_n} = 1$ nach sich zieht. Zu diesem Zweck ordnen wir die Elemente von Z in einer Folge an: $\lambda^{(1)}, \lambda^{(2)}, \ldots$. Nehmen wir an, die $d'_{\lambda^{(1)}}, \cdots, d'_{\lambda^{(m)}}$, d. h. die $\psi'_{\lambda^{(1)}}, \cdots, \psi'_{\lambda^{(m)}}$ seien schon wunschgemäß festgelegt, d. h. so, daß aus $k_1 \lambda^{(1)} + \cdots + k_m \lambda^{(m)} = 0$ ($k_1, \cdots, k_m = 0, \pm 1, \pm 2, \cdots$) $(\psi'_{\lambda^{(1)}})^{k_1} \cdot \ldots \cdot (\psi'_{\lambda^{(m)}})^{k_m} = 1$ folgt; wir wollen nun $d'_{\lambda^{(m+1)}}$, d. h. ψ'_{m+1}, so hinzudefinieren, daß auch aus $k_1 \lambda^{(1)} + \cdots + k_m \lambda^{(m)} + k_{m+1} \lambda^{(m+1)} = 0$ ($k_1, \cdots, k_m, k_{m+1} = 0, \pm 1, \pm 2, \cdots$)

$$(\psi'_{\lambda^{(1)}})^{k_1} \cdot \ldots \cdot (\psi'_{\lambda^{(m)}})^{k_m} \cdot (\psi'_{\lambda^{(m+1)}})^{k_{m+1}} = 1$$

folgt. Da die obige Forderung für $m = 0$ leer läuft, also erfüllt ist, können wir so induktiv alle $d'_{\lambda^{(1)}}, d'_{\lambda^{(2)}}, \cdots$ und $\psi'_{\lambda^{(1)}}, \psi'_{\lambda^{(2)}}, \cdots$, d. h. alle $d'_\lambda, \psi'_\lambda, \lambda$ von Z, unserer Bedingung entsprechend erzeugen. Betrachten wir also den Induktionsschritt von m zu $m+1$.

Betrachten wir die Gesamtheit der gültigen Gleichungen

$$k_1 \lambda^{(1)} + \cdots + k_m \lambda^{(m)} + k_{m+1} \lambda^{(m+1)} = 0 \quad (k_1, \cdots, k_m, k_{m+1} = 0, \pm 1, \pm 2, \cdots).$$

Wenn $k_{m+1} = 0$ ist, so ist

$$k_1 \lambda^{(1)} + \cdots + k_m \lambda^{(m)} = 0,$$

also nach Annahme

$$(\psi'_{\lambda^{(1)}})^{k_1} \cdot \ldots \cdot (\psi'_{\lambda^{(m)}})^{k_m} = 1,$$

$$\text{d. h. } (\psi'_{\lambda^{(1)}})^{k_1} \cdot \ldots \cdot (\psi'_{\lambda^{(m)}})^{k_m} (\psi'_{\lambda^{(m+1)}})^{k_{m+1}} = 1,$$

hier ist also, unabhängig von der Wahl von $d'_{\lambda^{(m+1)}}$, alles in Ordnung. Kommt also $k_{m+1} \neq 0$ gar nicht vor, so sind wir fertig. Kommt es vor, so sei Z_1 die Menge der k_{m+1}, die auftreten. Z_1 ist eine Menge ganzer Zahlen, es enthält offenbar mit k, l auch $k+l$, und es besteht nicht nur aus 0; somit ist Z_1 die Menge aller Vielfachen einer gewissen positiven ganzen Zahl $k^{(0)}$. Es gibt daher eine Gleichung

$$k_1^{(0)} \lambda^{(1)} + \cdots + k_m^{(0)} \lambda^{(m)} + k^{(0)} \lambda^{(m+1)} = 0.$$

Es genügt

$$(\psi'_{\lambda^{(1)}})^{k_1^{(0)}} \cdot \ldots \cdot (\psi'_{\lambda^{(m)}})^{k_m^{(0)}} (\psi'_{\lambda^{(m+1)}})^{k^{(0)}} = 1$$

zu sichern, denn für

$$k_1 \lambda^{(1)} + \cdots + k_m \lambda^{(m)} + k_{m+1} \lambda^{(m+1)} = 0$$

ist erstens $k_{m+1} = h k^{(0)}$ (h ganz), also

$$(k_1 - h k_1^{(0)}) \lambda^{(1)} + \cdots + (k_m - h k_m^{(0)}) \lambda^{(m)} = 0,$$

also nach Annahme

$$(\psi'_{\lambda^{(1)}})^{k_1-\hbar k_1^{(0)}} \cdot \ldots \cdot (\psi'_{\lambda^{(m)}})^{k_m-\hbar k_m^{(0)}} = 1,$$

und somit

$$(\psi'_{\lambda^{(1)}})^{k_1} \cdot \ldots \cdot (\psi'_{\lambda^{(m)}})^{k_m} \cdot (\psi'_{\lambda^{(m+1)}})^{k_{m+1}} = 1.$$

Nun ist wegen

$$\varphi'_{\lambda_1} \cdot \varphi'_{\lambda_2} = c'_{\lambda_1,\lambda_2}\, \varphi'_{\lambda_1+\lambda_2}, \quad |c'_{\lambda_1,\lambda_2}| = 1,$$

mit Beachtung von $\psi'_\lambda = d'_\lambda \varphi'_\lambda$, $|d'_\lambda| = 1$, jedenfalls

$$(\psi'_{\lambda^{(1)}})^{k_1^{(0)}} \cdot \ldots \cdot (\psi'_{\lambda^{(m)}})^{k_m^{(0)}} = b'_1\, \varphi'_{k_1^{(0)}\lambda^{(1)}+\ldots+k_m^{(0)}\lambda^{(m)}} = b'_1\, \varphi'_{-k^{(0)}\lambda^{(m+1)}},$$

$$(\varphi'_{\lambda^{(m+1)}})^{-k^{(0)}} = b'_2\, \varphi'_{-k^{(0)}\lambda^{(m+1)}}, \quad |b'_1| = |b'_2| = 1,$$

also

$$(\psi'_{\lambda^{(1)}})^{k_1^{(0)}} \cdot \ldots \cdot (\psi'_{\lambda^{(m)}})^{k_m^{(0)}} (\varphi'_{\lambda^{(m+1)}})^{k^{(0)}} = \frac{b'_1}{b'_2}.$$

Wir brauchen also bloß $d'_{\lambda^{(m+1)}}$ mit $(d'_{\lambda^{(m+1)}})^{k^{(0)}} = \dfrac{b'_2}{b'_1}$ zu wählen, und sind am Ziele ($|d'_{\lambda^{(m+1)}}|$ ist von selbst gleich 1).

Aus dem Bewiesenen folgt insbesondere

$$\psi'_{\lambda_1} \cdot \psi'_{\lambda_2} = \psi'_{\lambda_1+\lambda_2} \qquad (\lambda_3 = \lambda_1 + \lambda_2,\ k_1 = k_2 = 1,\ k_3 = -1).$$

Da wir an Stelle der ψ'_λ wieder φ'_λ schreiben können, dürfen wir annehmen, daß von vornherein $\varphi'_{\lambda_1} \cdot \varphi'_{\lambda_2} = \varphi'_{\lambda_1+\lambda_2}$ war. Ebenso zeigt man, daß wir annehmen dürfen, daß von vornherein $\varphi''_{\lambda_1} \cdot \varphi''_{\lambda_2} = \varphi''_{\lambda_1+\lambda_2}$ war.

Da die φ'_λ, λ in Z, und die φ''_λ, λ in Z, zwei vollständige normierte Orthogonalsysteme sind, wird durch

$$V\Big(\sum_{\lambda \text{ in } Z} a_\lambda \varphi'_\lambda\Big) = \sum_{\lambda \text{ in } Z} a_\lambda \varphi''_\lambda$$

ein unitärer Operator V definiert. Es ist $Vf \cdot Vg = V(f \cdot g)$ für $f = \varphi'_\lambda$, $g = \varphi'_\mu$, und hieraus schließt man, wie am Ende von 2., daß diese Gleichung für alle f, g gilt, wenn f beschränkt ist. D. h. V erfüllt D_5. Auf Grund des in § III, 1. Bewiesenen muß daher eine eineindeutige und maßtreue Abbildung Σ von Ω' minus einer Menge vom Maße 0 auf Ω'' minus eine Menge vom Maße 0 existieren, die V nach $Vf(P) = f(\Sigma P)$ erzeugt.

Ferner ist

$$U_t'' \varphi''_\lambda = e^{it\lambda} \varphi''_\lambda,\ V U_t' V^{-1} \varphi''_\lambda = V U_t' \varphi'_\lambda = V e^{it\lambda} \varphi'_\lambda = e^{it\lambda} \varphi''_\lambda,$$

d. h. es ist $U_t'' f = V U_t' V^{-1} f$

für alle $f = \varphi'_\lambda$. Daher gilt dies auch für die Linearaggregate der φ'_λ und deren Häufungspunkte, d. h. für alle f, also ist $U''_t = V U'_t V^{-1}$. Mit Hilfe der S'_t, S''_t, Σ ausgedrückt, besagt dies:

$$f(S''_t P'') = f(\Sigma S'_t \Sigma^{-1} P'') \ (P'' \text{ in } \Omega''),$$

und da dies für alle f gilt, muß $S''_t \cong \Sigma S'_t \Sigma^{-1}$ sein. D. h. die zwei allgemeinen Strömungen S'_t und S''_t sind isomorph.

4. Das bisher Bewiesene zusammenfassend, können wir also aussagen:

SATZ 5. *Wenn $\mu(\Omega) = 1$ und die allgemeine Strömung S_t in Ω (eigentlich-)ergodisch ist, und wenn die Schar U_t ein reines Punktspektrum hat, so ist dieses eine abzählbar-unendliche Menge Z von reellen Zahlen, die mit λ, μ auch $\lambda \pm \mu$ enthält.*

Umgekehrt werden auf diese Weise alle abzählbar-unendlichen Mengen Z von reellen Zahlen, die mit λ, μ auch $\lambda \pm \mu$ enthalten, erzeugt, und zwar bestimmt Z Ω, S_t im wesentlichen (d. h. bis auf isomorphe Abbildungen Σ, vgl. D_3) eindeutig.

Jede Zahlenmenge Z der beschriebenen Art, d. h. jeder „Zahlenmodul", bestimmt somit im wesentlichen eindeutig eine eigentlich-ergodische Strömung. Damit ist die Existenz unendlich vieler Typen ergodischer Strömungen bewiesen.

Indessen ist es lehrreich, für die einfachsten „Zahlenmoduln" Z Beispiele dazugehöriger Strömungen (die ja im wesentlichen die einzigen sind) effektiv anzugeben.

Wenn z. B. Z eine endliche Basis hat — d. h. wenn ein Zahlensystem $\omega_1, \cdots, \omega_n$ existiert, so daß erstens keine Relation $k_1\omega_1 + \cdots + k_n\omega_n = 0$ $(k_1, \cdots, k_n = 0, \pm 1, \pm 2, \cdots)$, außer mit $k_1 = \cdots = k_n = 0$, besteht, und zweitens Z die Menge aller $k_1\omega_1 + \cdots + k_n\omega_n$ $(k_1, \cdots, k_n = 0, \pm 1, \pm 2, \cdots)$ ist — so werden wir zu einer wohlbekannten, von Weyl entdeckten Strömung geführt[61].

Sei nämlich Ω der n-dimensionale Würfel, $P = [u_1, \cdots, u_n]$, $0 \leqq u_1 < 1, \cdots, 0 \leqq u_n < 1$, μ^* das gewöhnliche Lebesguesche äußere Maß. Es sei

$$S_t[u_1, \cdots, u_n] = \left[\left\{u_1 + \frac{\omega_1}{2\pi} t\right\}, \cdots, \left\{u_n + \frac{\omega_n}{2\pi} t\right\}\right]$$

($\{u\}$ ist der Rest von u mod 1: $0 \leqq \{u\} < 1$, $u - \{u\}$ ganz). Die Funktionen $e^{2\pi i(k_1 u_1 + \cdots + k_n u_n)}$ $(k_1, \cdots, k_n = 0, \pm 1, \pm 2, \cdots)$ bilden bekanntlich ein vollständiges normiertes Orthogonalsystem; und wegen

$$S_t e^{2\pi i(k_1 u_1 + \cdots + k_n u_n)} = e^{i(k_1\omega_1 + \cdots + k_n\omega_n)t} e^{2\pi i(k_1 u_1 + \cdots + k_n u_n)}$$

[61] H. Weyl, Math. Ann. 77, 1916, S. 313. Als Beispiel einer ergodischen Strömung auch bei E. Hopf, a. a. O. Anm. [3].

sind es Eigenfunktionen mit den bzw. Eigenwerten $k_1\omega_1+\cdots+k_n\omega_n$. Also gibt es keine weiteren Eigenfunktionen. Somit liegt ein reines Punktspektrum vor, das gleich Z ist; und 0 ist einfacher Eigenwert mit der Eigenfunktion 1, d. h. die Strömung ist eigentlich-ergodisch. Dieses Ω und S_t ist also das gesuchte Beispiel.

Es sei noch erwähnt, daß wir hier für $n=1$ die Menge der $k\omega$ ($k=0,\pm1,\pm2,\cdots$) als Z erhalten, und S_t die einfach periodische Bewegung mit der Periode $\frac{2\pi}{\omega}$ wird. Dieses Z ist bekanntlich der einzige nicht überall dichte „Zahlmodul“ [62], in allen anderen Fällen mit reinem Punktspektrum ist also dieses überall dicht.

Ein etwas komplizierterer Fall ist dieser: Z ist die Menge aller dyadisch rationalen Zahlen $\left(\frac{n}{2^p},\ n=0,\pm1,\pm2,\cdots,\ p=1,2,\cdots\right)$, denn hier existiert keine Basis. Ω sei das Quadrat $P=[u,v]$, $0\leqq u<1$, $0<v<1$, μ^* sei das gewöhnliche Lebesguesche äußere Maß. S_t werde so definiert: $\varphi(v)$ ist die folgende eineindeutige und maßtreue Abbildung von $0<v<1$ auf sich selbst: $\frac12<v<1$ wird um $-\frac12$ nach $0<\varphi(v)<\frac12$ verschoben, $\frac14<v<\frac12$ um $+\frac14$ nach $\frac12<\varphi(v)<\frac34$, $\frac18<v<\frac14$ um $+\frac58$ nach $\frac34<v<\frac78$, $\cdots$, die übrigbleibenden Punkte $v=\frac12,\frac14,\frac18,\cdots$ werden in dieser Reihenfolge den übrigbleibenden $\varphi(v)=\frac12,\frac34,\frac78,\cdots$ zugeordnet [63]. $S_t[u,v]$ entsteht, indem man u um $\frac{t}{2\pi}$ wachsen läßt (d. h. für $t>0$ nach rechts, für $t<0$ nach links verschiebt), aber so oft man hierbei den rechten oder linken Rand von Ω (d. i. $u=1$ bzw. $u=0$) passiert, u auf den linken bzw. rechten Rand zurückversetzt und v durch $\varphi(v)$ bzw. $\varphi^{-1}(v)$ (Inverse!) ersetzt [64]. Das Spektrum der U_t-Schar bestimmt man am bequemsten so: Man versuche eine in jedem der 2^p Intervalle $\frac{m-1}{2^p}<v<\frac{m}{2^p}$

---

[62] Z hat Elemente $\lambda\neq0$, also auch Elemente >0: $|\lambda|=\pm\lambda$. Gibt es unter seinen Elementen >0 ein kleinstes, λ_0, so schließen wir so: für jedes λ von Z gibt es ein $k=0,\pm1,\pm2,\cdots$ mit $\lambda-k\lambda_0\geqq0,<\lambda_0$, also kann nicht $\lambda-k\lambda_0>0$ sein, also ist $\lambda-k\lambda_0=0$, $\lambda=k\lambda_0$. D. h. Z ist die im Text genannte Menge. Andernfalls wird die untere Grenze der $\lambda>0$ von Z, λ_0, nicht angenommen. Zu $\varepsilon>0$ gibt es ein λ_1 in Z mit $\lambda_0<\lambda_1<\lambda_0+\varepsilon$, und ein λ_2 in Z mit $\lambda_0<\lambda_2<\lambda_1$. Also ist das $\lambda=\lambda_1-\lambda_2$ von Z >0, $<\varepsilon$. Bei beliebigem μ können wir daher $\mu<k\lambda<\mu+\varepsilon$ erreichen, da μ, $\varepsilon>0$ beliebig sind, ist Z überall dicht.

[63] In Formeln: $\varphi(v)=v+1-\frac{3}{2^n}$ für $\frac{1}{2^n}<v<\frac{1}{2^{n-1}}$, und $\varphi(v)=1-v$ für $v=\frac{1}{2^n}$ ($n=1,2,\cdots$).

[64] In Formeln: $\varphi^n(v)$ sei die $|n|$-fache Iterierte von $\varphi(v)$ bzw. $\varphi^{-1}(v)$ (für $n>$bzw.<0), dann ist $S_t[u,v]=\left[\left\{u+\frac{t}{2\pi}\right\},\ \varphi^{u+\frac{t}{2\pi}-\left\{u+\frac{t}{2\pi}\right\}}(v)\right]$ $\left(u+\frac{t}{2\pi}-\left\{u+\frac{t}{2\pi}\right\}\text{ ist ganz}\right)$.

$(m = 1, 2, \cdots, 2^p)$ konstante Funktion $\sigma_{n,p}(v)$ so zu bestimmen, daß $\varphi_{n,p}(u, v) = \sigma_{n,p}(v)\, e^{2\pi i \frac{n}{2^p} u}$ eine normierte Eigenfunktion vom Eigenwert $\frac{n}{2^p}$ wird. Es zeigt sich, daß dies für jedes $n = 0, \pm 1, \pm 2, \cdots$, $p = 1, 2, \cdots$, auf genau eine Weise gelingt (bis auf einen unwesentlichen konstanten Faktor). Da $\varphi_{n,p}$ den an $\varphi_{2n,p+1}$ gestellten Anforderungen genügt, ist $\varphi_{n,p} = \varphi_{2n,p+1}$; also $\varphi_{n,p}$ nur vom Werte von $\lambda = \frac{n}{2^p}$ abhängig, also auch mit φ_λ bezeichenbar. Da sie Eigenfunktionen sind, bilden die φ_λ ein normiertes Orthogonalsystem, es zeigt sich, daß es ein vollständiges ist. Also gibt es keine weiteren Eigenfunktionen. Somit liegt ein reines Punktspektrum vor, das gleich Z ist; und 0 ist einfacher Eigenwert, natürlich mit der Eigenfunktion 1, d. h. die Strömung ist eigentlich-ergodisch. Ω, S_t sind also das gesuchte Beispiel. Die einfache Durchführung der zwei weiter oben ausgelassenen Details übergehen wir.

Nach ähnlichen Schemata, wie diese zwei, ließen sich noch viele andere „Zahlmoduln“ Z behandeln, worauf wir nicht mehr eingehen.

V. Bemerkungen über U_t-Scharen mit Streckenspektrum.

1. Die U_t-Scharen mit reinem Streckenspektrum (d. h. vom eventuellen einfachen Eigenwert 0 abgesehen, ohne Punktspektrum) sind, wie schon in Einleitung 4 betont wurde, aus zwei Gründen von besonderer Wichtigkeit. Erstens gilt die Verallgemeinerung des Ergodensatzes (e) in Einleitung 2, vgl. a. a. O. Anm.[3]) für sie und nur für sie, d. h. sie sind, wie a. a. O. des näheren erläutert wurde, die Strömungen ohne „à la longue“ gültige Eigenschaften. Zweitens ist zu vermuten, daß die U_t-Scharen im allgemeinen ein reines Streckenspektrum haben, d. h. daß das Auftreten eines Punktspektrums eine seltene Ausnahmeerscheinung ist. Dies letztere werden wir im § VI plausibel zu machen suchen. Auf Grund dieser Umstände wäre es erwünscht, sie mindestens so gut zu übersehen, wie dies in § IV für die Punktspektra gelang.

Zunächst wäre zu ermitteln, ob nicht das Streckenspektrum, wie das Punktspektrum, stets einfach ist (ergodische Strömung vorausgesetzt!). Weiter ist es wahrscheinlich, daß ein Isomorphiesatz gilt, wie er in Einleitung 3. formuliert und in der zweiten Hälfte von Satz 5 für das reine Punktspektrum bewiesen wurde. D. h. daß zwei allgemeine Strömungen Ω, S_t mit unitär ineinander transformierbaren U_t einander im wesentlichen (d. h. bis auf eine eineindeutige und maßtreue Abbildung) gleich sind. Dann wüßten wir, was alle für uns bedeutungsvollen Eigenschaften einer Strömung (d. s. die Isomorphie-Invarianten, vgl. Einleitung 3.) festlegt: so wie dies im Falle des Punkt-

spektrums nach Satz 5 das Punktspektrum Z war, wäre es im Falle des Streckenspektrums dieses und sein Hellingerscher Typus (vg. Anm. [25]). Hinzu tritt noch, daß wir, wenigstens für differentiierbare Strömungen, zeigen werden (und es ist für alle zu vermuten): wenn ein Streckenspektrum überhaupt existiert, so umfaßt es die ganze Zahlengerade $-\infty<\lambda<+\infty$. Also käme es nur auf den Hellingerschen Typus an.

Schließlich wissen wir aus Satz 5, daß das Punktspektrum Z nicht jede beliebige (abzählbar-unendliche) Zahlenmenge sein kann: vielmehr ist die Beschränkung da, daß Z ein „Zahlmodul" sein muß, aber keine weitere. Es wäre also wichtig, zu erfahren, welchen Beschränkungen der Hellingersche Typus unterliegt. Wenn er z. B. immer der einfachste Typus („λ") wäre, so würde dies eine bemerkenswerte Verschärfung von e) (Einleitung 2., vgl. die zweite in Anm. [3] genannte Abhandlung, ferner Anm. [25]) nach sich ziehen. Ferner hätte es die überraschende Folge, daß alle Strömungen mit reinem Streckenspektrum (d. h. gerade der vermutlich allgemeine Fall) einander im wesentlichen gleich (d. h. isomorph) wären. Indessen ist dies nicht sehr wahrscheinlich.

Von alledem können wir leider einstweilen nichts beweisen. Bloß die Ausdehnung des Streckenspektrums bei differentiierbaren Strömungen soll im Folgenden bestimmt werden.

2. Ω sei ein m-Raum, der „im kleinen" auf n-dimensionale Koordinaten bezogen werden kann (d. h. in einer geeigneten Umgebung eines jeden P_0 $P=[u_1,\cdots,u_n]$), und S_t eine ergodische und differentiierbare Strömung in Ω.

Wenn P ein Punkt von Ω ist, so sei $S(P)$ die Bahnkurve S_tP, $-\infty<t<+\infty$. Wenn S_tP Doppelpunkte hat, d. h. $S_{t_1}P=S_{t_2}P$, $t_1<t_2$, vorkommt, so ergibt das Anwenden von S_{t-t_1}: $S_tP=S_{t+(t_2-t_1)}P$, d. h. es gibt eine positive Periode t_2-t_1, die kleinste solche Periode sei τ. Tritt dies für jedes P ein, so ist $\tau=\tau(P)>0$, wie man leicht einsieht, eine meßbare Funktion von P. Es gibt also ein P_0, wo sie approximativ stetig ist (dies ist sogar für alle P_0 bis auf eine Menge vom Maße 0 der Fall[65]). Sei K eine Umgebung von P_0, K_1 die Menge der S_tP, P in K, $0\leqq t\leqq\tau(P_0)+1$. Das Maß dieser Menge ist offenbar endlich, und da sie bei sich auf P_0 zusammenziehendem K auf den Bahnkurvenbogen S_tP_0, $0\leqq t\leqq\tau(P_0)+1$, d. h. $S(P_0)$, zusammenzieht (weil S_tP stetig ist!), strebt das Maß gegen das Maß von $S(P_0)$. Ist dieses $<\mu(\Omega)$, so können wir K mit $\mu(K_1)<\mu(\Omega)$ wählen. Da $\tau(P)$ bei $P=P_0$ approximativ stetig ist, ist in einer Menge $A\subseteq K$, $\mu(A)>0$, $\tau(P)\leqq\tau(P_0)+1$ (es ließe sich noch mehr behaupten). Sei A_1 die Menge der S_tP, P in A, $0\leqq t\leqq\tau(P_0)+1$, wegen $\tau(P)\leqq\tau(P_0)+1$ ist A_1 die Summe der $S(P)$, P in A, also S_t-invariant. Dabei ist

[65] Vgl. z. B. Sierpinski, Fund. Math. 3, 1922, S. 320.

$$\mu(\mathcal{A}_1) \geqq \mu(\mathcal{A}) > 0, \qquad \leqq \mu(K_1) < \mu(\Omega).$$

Somit wäre S_t nicht ergodisch.

Also muß $\mu(S(P_0)) = \mu(\Omega)$ sein. Da wir eine Menge vom Maße 0 fortlassen können, können wir $S(P_0) = \Omega$ annehmen. $S(P_0)$ ist eineindeutig auf die $s \geqq 0$, $< \tau (= \tau(P_0))$ bezogen (durch $s \to S_s P_0$). Wir können Ω somit durch die Menge $0 \leqq t < \tau$ ersetzen unter Übertragung des Maßes. Da

$$S_t S_s = S_{t+s} = S_{t+s-e\tau} \qquad (e = 0, \pm 1, \pm 2, \cdots)$$

ist, ist nunmehr S_t durch $S_t s = t + s \cdots \bmod \tau$ definiert. Daraus, daß das Maß dieser Operation gegenüber invariant sein muß, folgert man unschwer, daß es das Lebesguesche ist, eventl. mit einer Konstanten $a > 0$ multipliziert. Und nunmehr sieht man, daß die Funktionen $\frac{1}{\sqrt{a\tau}} e^{\frac{2\pi i}{\tau} ks}$, $k = 0, \pm 1, \pm 2, \cdots$, Eigenfunktionen der U_t sind, mit den bzw. Eigenwerten $\frac{2\pi}{\tau} k$; da sie ein vollständiges normiertes Orthogonalsystem bilden, gibt es keine anderen Eigenfunktionen. Es liegt also ein reines Punktspektrum vor, Z: $\frac{2\pi}{\tau} k$, $k = 0, \pm 1, \pm 2, \cdots$. Das ist gerade der in § IV, 4. diskutierte allereinfachste Fall: einfach periodische Bewegung. Insbesondere ist $n = 1$.

Wenn dieser nicht vorliegt, so muß es demnach ein P_0 mit einfach durchlaufener Bahnkurve $S(P_0)$ geben: alle $S_t P_0$ voneinander verschieden. Wir wählen nun ein t_0, das wir vorläufig festhalten. Ein quer zu $S(P_0)$ liegendes, $n-1$-dimensionales Flächenelement durch P_0 in Ω, F, können wir so wählen, daß alle $S_t F$, $0 \leqq t \leqq t_0$ zueinander fremd sind (Differentiierbarkeit!, F muß nur klein genug sein). Sei $F(t_1, t_2)$ $(t_1 < t_2)$ die Menge der $S_t P$, P in F, $t_1 \leqq t < t_2$. Da $F(t_1, t_2)$ ein n-dimensionaler Bereich ist, ist $\mu(F(t_1, t_2))$ stets > 0. Aus der Maßtreue von S_t folgt

$$\mu(F(t_1, t_2)) = \mu(S_t F(t_1, t_2)) = \mu(F(t_1 + t, t_2 + t)),$$

d. h.

$$\mu(F(t_1, t_2)) = f(t_2 - t_1).$$

Liegen t_1, t_2, t_3 in $0 \leqq t \leqq t_0$, so ist

$$\mu(F(t_1, t_2)) + \mu(F(t_2, t_3)) = \mu(F(t_1, t_3)),$$

d. h. es ist

$$f(s') + f(s'') = f(s' + s''),$$

wenn $s', s'' > 0$, $s' + s'' \leqq t_0$ ist. Da ferner $f(s)$ stets > 0 ist (es ist ja nur für $s > 0$, $\leqq t_0$ definiert), erkennt man, daß $f(s) = cs$, $c =$ Konstans > 0, sein muß. Also $\mu(F(t_1, t_2)) = c(t_1 - t_2)$.

Wir definieren nun: $\varphi(P) \begin{cases} = e^{i\lambda s}, & \text{wenn } P \text{ in } S_s F \text{ liegt, } 0 \leqq s < t_0 \\ = 0, & \text{sonst (d. h. } P \text{ nicht in } F(0, t_0)) \end{cases}$.

Es ist

$$\|\varphi\|^2 = \int_{F(0,t_0)} ds = \mu(F(0, t_0)) = ct_0,$$

$$\|U_t\varphi_0 - e^{i\lambda t}\varphi_0\|^2 = \int_{F(0,t)+F(t_0,t_0+t)} ds = \mu(F(0, t)) + \mu(F(t_0, t_0+t)) = 2ct,$$

für $0 \leqq t \leqq t_0$. Nun ist $(E(\lambda) - E(-\lambda))\varphi \to \varphi$ für $\lambda \to \infty$, also für ein geeignetes λ

$$\|(E(\lambda) - E(-\lambda))\varphi\| \geqq \frac{1}{2} ct_0.$$

Andererseits ist $E(\lambda) - E(-\lambda)$ ein Projektionsoperator, und mit U_t vertauschbar, also

$$\begin{aligned} &\|U_t(E(\lambda) - E(-\lambda))\varphi - e^{it\lambda}(E(\lambda) - E(-\lambda))\varphi\| \\ &= \|(E(\lambda) - E(-\lambda))(U_t\varphi - e^{it\lambda}\varphi)\| \leqq \|U_t\varphi - e^{it\lambda}\varphi\| = 2ct. \end{aligned}$$

Für $\psi = \dfrac{(E(\lambda) - E(-\lambda))\varphi}{\|(E(\lambda) - E(-\lambda))\varphi\|}$ haben wir also:

$$\|\psi\| = 1, \qquad \|U_t\psi - e^{it\lambda}\psi\| \leqq \frac{4t}{t_0}.$$

Ferner ist A auf $(E(\lambda) - E(-\lambda))\varphi$ anwendbar[66], also auch auf ψ. Unsere Ungleichheit kann auch so geschrieben werden:

$$\left\| \frac{U_t - 1}{it}\psi - \frac{e^{i\lambda t} - 1}{it}\psi \right\| \leqq \frac{4}{t_0},$$

und hieraus folgt durch $t \to 0$:

$$\|A\psi - \lambda\psi\| \leqq \frac{4}{t_0}.$$

t_0 war beliebig, wir können es also, wenn ein beliebiges $\varepsilon > 0$ gegeben ist, mit $\dfrac{4}{t_0} < \varepsilon$ wählen. Dann wird $\|A\psi - \lambda\psi\| < \varepsilon\|\psi\|$. Da für jedes $\varepsilon > 0$ ein solches ψ gefunden werden kann, gehört λ zum Gesamtspektrum. D. h.: das Gesamtspektrum ist die ganze Zahlengerade $-\infty < \lambda < +\infty$.

3. Wenn kein Punktspektrum da ist, oder das Punktspektrum Z die Menge der $k\omega$, $k = 0, \pm 1, \pm 2, \cdots$, für ein festes $\omega > 0$ ist, so enthält das Streckenspektrum alle Punkte des Gesamtspektrums, bis auf höchstens

[66] Vgl. z. B. A, S. 415, Anm. [68].

die $k\omega$. Wenn also nicht der in 2. erwähnte Spezialfall (reines Punktspektrum, $Z =$ Menge der $k\omega$, $k = 0, \pm 1, \pm 2, \cdots$) vorliegt, so enthält das Gesamtspektrum alle λ, so daß das Streckenspektrum überall dicht ist. Da es eine abgeschlossene Menge ist, enthält es alle λ.

Andernfalls ist das Punktspektrum Z überall dicht (vgl. Anm. [62]). Gehöre λ zu Z, φ_λ sei eine Eigenfunktion vom Eigenwert λ. Nach Satz 4 ist $|\varphi_\lambda| =$ Konstans > 0, wir können also $|\varphi_\lambda| = 1$ annehmen. Der Operator $Vf = \varphi_\lambda \cdot f$ ist also unitär. Aus $U_t \varphi_\lambda = e^{it\lambda} \varphi_\lambda$ folgt

$$V^{-1} U V = e^{it\lambda} U.$$

Nun hat $V^{-1} U V$ dasselbe Spektrum wie U, $e^{it\lambda} U$ dagegen offenbar ein um λ nach rechts verschobenes. Insbesondere das Streckenspektrum von U geht also durch eine Verschiebung nach rechts um λ in sich selbst über. Wenn also μ dazu gehört, gehören alle $\mu + \lambda$, λ in Z, auch dazu, es ist also überall dicht. Da es eine abgeschlossene Menge ist, enthält es dann alle μ. D. h.: es ist leer, oder die ganze Zahlengerade $-\infty < \lambda < +\infty$.

Zusammenfassend können wir also sagen:

SATZ 5. *Für eine ergodische differentiierbare Strömung S_t ist das Gesamtspektrum immer die ganze Zahlengerade $-\infty < \lambda < +\infty$, mit einer Ausnahme: bei der einfach periodischen Strömung mit der Periode $\frac{2\pi}{\omega}$, $\omega > 0$, und dem reinen Punktspektrum Z: $k\omega$, $k = 0, \pm 1, \pm 2, \cdots$.*

Das Streckenspektrum fehlt, oder es ist die ganze Zahlengerade $-\infty < \lambda < +\infty$.

Die genaue Betrachtung des Beweises lehrt, daß die Differentiierbarkeits-Annahme abgeschwächt werden kann: z. B. schaden endlich viele Ausnahmestellen in Ω nichts. Wahrscheinlich gilt der Satz aber auch unabhängig von jeder Annahme über die Regularität der Strömung.

VI. Beispiele für ein reines Streckenspektrum.

1. Aus § IV, 1. wissen wir, daß jede Null-ergodische Strömung ein reines Streckenspektrum hat. Solche Strömungen sind leicht anzugeben: z. B. Ω die Menge aller reellen u, μ^* das gewöhnliche Lebesguesche äußere Maß, $S_t u = u + t$. Denn wenn A meßbar und von endlichem Maße ist, so können wir $t_0 > 0$ so wählen, daß der in $-t_0 < u < t_0$ liegende Teil von A ein Maß $> \frac{1}{2}\mu(A)$ hat. Der in $-3t_0 < u < -t_0$ liegende hat daher ein Maß $< \frac{1}{2}\mu(A)$, also auch der in $-t_0 < u < t_0$ liegende Teil von $S_{2t_0} A$, so daß sich $A, S_{2t_0} A$ nicht nur um eine Menge vom Maße 0 unterscheiden können. Jede A-Menge muß daher das Maß 0 haben, d. h. S_t ist Null-ergodisch.

Uns interessieren aber in erster Linie die eigentlich-ergodischen Strömungen, und daher wollen wir für solche Beispiele mit reinem Streckenspektrum angeben.

Die zu diskutierenden Beispiele werden alle von dieser Art sein: $f(u)$ sei eine in $0 \leqq u < 1$ definierte Funktion, die stets $\geqq A, \leqq B$ bleibt, für zwei feste positive A, B; γ eine Zahl, $0 \leqq \gamma < 1$. Ω sei der ebene Bereich $P = [u, v]$, $0 \leqq u < 1$, $0 \leqq v < f(u)$, μ^* das gewöhnliche Lebesguesche äußere Maß[67]. $S_t[u, v]$ entsteht, indem man v um t wachsen läßt (für $t > 0$, für $t < 0$ ist das eine Abnahme), aber jedesmal, wenn man die obere oder die untere Grenze von Ω ($v = f(u)$ bzw. $v = 0$) passiert, v auf die untere bzw. obere Grenze von Ω zurückversetzt, und u durch $\{u+\gamma\}$ bzw. $\{u-\gamma\}$ ersetzt (d. h. durch diejenige der Zahlen $u+\gamma$, $u+\gamma-1$ bzw. $u-\gamma$, $u-\gamma+1$, die $\geqq 0$, < 1 ist). S_t ist offenbar maßtreu.

Wir werden nun untersuchen, für welche $f(u), \gamma$ diese Strömung (eigentlich-) ergodisch ist, und ein reines Streckenspektrum hat.

2. Gehöre λ zum Punktspektrum der zu den S_t gehörigen Operatorenschar U_t, $\varphi \neq 0$ sei eine dazugehörige Eigenfunktion. D. h. es soll für jedes feste t, bis auf eine $[u, v]$-Menge vom Maße 0, $\varphi(S_t[u, v]) = e^{i\lambda t}\varphi(u, v)$ sein. Wir werden nun mit Hilfe des Umstandes, daß S_t nicht nur D_2', sondern auch D_2 erfüllt, die obigen Ausnahmemengen eliminieren.

Da $\varphi(u, v)$ in $[u, v]$ meßbar ist, und das $S_t[u, v]$-Urbild jeder meßbaren Menge eine meßbare $[u, v, t]$-Menge ist (vgl. § II, 6), ist $\varphi(S_t[u, v])$ in $[u, v, t]$ meßbar. Die $[u, v, t]$-Menge $\bar{N}$, die durch $\varphi(S_t[u, v]) \neq e^{i\lambda t}\varphi(u, v)$ definiert ist, ist also meßbar. Für jedes feste t ist die Menge $\bar{N}_{(t)}$ der $[u, v]$ mit $[u, v, t]$ in $\bar{N}$ eine Menge vom Maße 0. Nach dem Satz von Fubini ist also auch $\bar{N}$ vom Maße 0. Nach demselben Satz (in umgekehrter Richtung angewandt), ist für alle $[u, v]$, abgesehen von einer $[u, v]$-Menge N vom Maße 0, die Menge der t mit $[u, v, t]$ in $\bar{N}$ vom Maße 0. D. h.: wenn $[u, v]$ nicht zu N gehört, gilt $\varphi(S_t[u, v]) = e^{i\lambda t}\varphi(u, v)$ für alle t, bis auf eine Menge vom Maße 0.

Nun sollen $[u', v']$, $[u'', v'']$ nicht zu N gehören, ferner sei

$$S_{t'}[u', v'] = S_{t''}[u'', v''].$$

Es gilt

$$\varphi(S_{s'}[u', v']) = e^{i\lambda s'}\varphi(u', v'), \qquad \varphi(S_{s''}[u'', v'']) = e^{i\lambda s''}\varphi(u'', v''),$$

ferner

$$\varphi(S_t S_{t'}[u', v']) = e^{i\lambda t}\varphi(S_{t'}[u', v']), \qquad \varphi(S_t S_{t''}[u'', v'']) = e^{i\lambda t}\varphi(S_{t''}[u'', v'']),$$

bis auf je eine Menge vom Maße 0, bzw. in s', s'', t, t. Binden wir s', s'', t durch $s' = t+t'$, $s'' = t+t''$ aneinander, so gelten alle vier Gleichungen bis auf eine t-Menge vom Maße 0. Wählen wir also t so,

[67] Oder, für ein festes positives c, das c-fache davon. Dies würde am folgenden nichts ändern, und wir könnten es benützen, um $\mu(\Omega) = c\int_0^1 f(u)\,du$ gleich 1 zu machen.

daß sie gelten, und beachten $S_t S_{t'} = S_{s'}$, $S_t S_{t''} = S_{s''}$ sowie $S_{t'}[u', v'] = S_{t''}[u'', v'']$, so wird

$$e^{i\lambda s'} \varphi(u', v') = e^{i\lambda s''} \varphi(u'', v''), \quad e^{i\lambda t'} \varphi(u', v') = e^{i\lambda t''} \varphi(u'', v'').$$

Wenn wir also für alle $[\bar{u}, \bar{v}] = S_t[u, v]$, $[u, v]$ nicht in N, ein neues $\bar{\varphi}(\bar{u}, \bar{v})$ durch $e^{i\lambda t} \varphi(u, v)$ definieren, so ist diese Definition nicht selbstwidersprechend. Für die übrigen $[\bar{u}, \bar{v}]$ sei etwa $\bar{\varphi}(\bar{u}, \bar{v}) = 0$. Gehört $[\bar{u}, \bar{v}]$ nicht zu N, so können wir $u = \bar{u}$, $v = \bar{v}$, $t = 0$ setzen, und es wird $\bar{\varphi}(\bar{u}, \bar{v}) = \varphi(\bar{u}, \bar{v})$. Also tritt $\bar{\varphi}(u, v) \neq \varphi(u, v)$ nur in N, also in eine Menge vom Maße 0 ein, d. h. $\bar{\varphi}$ gehört, wie φ, zu $\mathfrak{H}$, ja es ist als $\mathfrak{H}$-Element gleich φ. Ferner gilt nach Konstruktion ausnahmslos

$$\bar{\varphi}(S_t[u, v]) = e^{i\lambda t} \bar{\varphi}(u, v).$$

Anstatt φ durch $\bar{\varphi}$ zu ersetzen, können wir auch annehmen, daß φ von vornherein so war.

3. Wenn $[u, v']$, $[u, v'']$ beide in Ω liegen, so ist für $t = v'' - v'$ $S_t[u, v'] = [u, v'']$, also

$$\varphi(u, v'') = e^{i\lambda(v''-v')} \varphi(u, v'), \quad e^{-i\lambda v'} \varphi(u, v') = e^{-i\lambda v''} \varphi(u, v'').$$

$e^{-i\lambda v} \varphi(u, v)$ ist also von v unabhängig, es ist somit eine meßbare u-Funktion: $\psi(u)$. Sei nun $0 < t < 2A$, $v = f(u) - \frac{t}{2}$, dann ist

$$S_t\left[u, f(u) - \frac{t}{2}\right] = \left[\{u+\gamma\}, \frac{t}{2}\right],$$

also

$$e^{i\lambda\frac{t}{2}} \psi(\{u+\gamma\}) = e^{i\lambda t} \cdot e^{i\lambda\left(f(u)-\frac{t}{2}\right)} \psi(u), \quad \psi(\{u+\gamma\}) = e^{i\lambda f(u)} \psi(u).$$

Umgekehrt erkennt man: wenn $\psi(u)$ meßbar ist und

$$\psi(\{u+\gamma\}) = e^{i\lambda f(u)} \psi(u)$$

gilt, so ist

$$\varphi(u, v) = e^{i\lambda v} \psi(u)$$

eine Lösung von

$$\varphi(S_t[u, v]) = e^{i\lambda t} \varphi(u, v).$$

Da ferner

$$\iint_\Omega |\varphi(u, v)|^2 \, du \, dv = \int_0^1 \int_0^{f(u)} |\psi(u)|^2 \, du \, dv$$

$$= \int_0^1 |\psi(u)|^2 f(u) \, du \begin{cases} \geqq A \cdot \int_0^1 |\psi(u)|^2 \, du \\ \leqq B \cdot \int_0^1 |\psi(u)|^2 \, du \end{cases}$$

ist, ist die Erfüllbarkeit der obigen Bedingungen für $\psi(u)$, zusammen mit der Endlichkeit von $\int_0^1 |\psi(u)|^2 du$, für den Eigenwert-Charakter von λ charakteristisch.

Für $\lambda = 0$ besagt dies $\psi(\{u+\gamma\}) = \psi(u)$. Nun bilden die $e^{2\pi iku}$, $k = 0, \pm 1, \pm 2, \cdots$, ein vollständiges normiertes Orthogonalsystem, und sind Eigenfunktionen des unitären Operators $U\chi(u) = \chi(\{u+\gamma\})$, und zwar mit den bzw. Eigenwerten $e^{2\pi ik\gamma}$. Also gibt es keine weiteren Eigenfunktionen. Ist γ rational, so sind viele $e^{2\pi ik\gamma} = 1$, also viele $\psi(u) = e^{2\pi iku}$ möglich, d. h. unsere Strömung nicht ergodisch. Ist es dagegen irrational, was im folgenden vorausgesetzt werde, so folgt aus $e^{2\pi ik\gamma} = 1$ $k = 0$, also muß $\psi(u)$ ein konstantes Vielfaches der einzigen Eigenfunktion, 1, sein — also $\psi(u) = c$, $\varphi(u, v) = c$, d. h. unsere Strömung ist ergodisch.

Sei nun $\lambda \neq 0$. Es ist $|\psi(\{u+\gamma\})| = |\psi(u)|$, also die obigen Schlüsse auf $|\psi(u)|$ anwendbar: es ist $|\psi(u)| = c$, wegen $\varphi \neq 0$, $\psi \neq 0$ ist $c > 0$. $\int_0^1 |\chi(\{u+\varepsilon\}) - \chi(u)|^2 du \to 0$ für $\varepsilon \to 0$ ist eine wohlbekannte Tatsache[68]. — Nach dem Satze von Kronecker[69] können wir, da γ irrational ist, eine Folge ganzer Zahlen $k_1 < k_2 < \cdots$ mit $\{k_n\gamma\} \to 0$ für $n \to \infty$ finden. Also ist dann

$$\int_0^1 |\psi(\{u+k_n\gamma\}) - \psi(u)|^2 du = \int_0^1 |\psi(\{u+\{k_n\gamma\}\}) - \psi(u)|^2 du \to 0.$$

Nun ist andererseits

$$\psi(\{u+k\gamma\}) = e^{i\lambda(f(u)+f(\{u+\gamma\})+\cdots+f(\{u+(k-1)\gamma\}))} \cdot \psi(u),$$

also

$$\begin{aligned}
&\int_0^1 |\psi(\{u+k\gamma\}) - \psi(u)|^2 du \\
&= \int_0^1 |(e^{i\lambda(f(u)+f(\{u+\gamma\})+\cdots+f(\{u+(k-1)\gamma\}))} - 1)\psi(u)|^2 du \\
&= c^2 \int_0^1 |e^{i\lambda(f(u)+f(\{u+\gamma\})+\cdots+f(\{u+(k-1)\gamma\}))} - 1|^2 du \\
&= 4c^2 \int_0^1 \sin^2\left(\frac{\lambda}{2}(f(u)+f(\{u+\gamma\})+\cdots+f(\{u+(k-1)\gamma\}))\right) du.
\end{aligned}$$

Wenn wir also

$$\operatorname*{limes\,inf}_{k\to\infty} \left[\int_0^1 \sin^2\left(\frac{\lambda}{2}(f(u)+f(\{u+\gamma\})+\cdots+f(\{u+(k-1)\gamma\}))\right) du\right] > 0$$

[68] Es ist evident für alle $\chi(u) = e^{2\pi iku}$, $k = 0, \pm 1, \pm 2, \cdots$, also gilt es auch für deren Linearaggregate und für die Häufungspunkte der Linearaggregate — also für alle $\chi(u)$.

[69] Vgl. z. B. bei Weyl, a. a. O. Anm. [54].

beweisen können, so ist die Unerfüllbarkeit der $\psi(u)$-Bedingung, d. h. die Nichtexistenz des Punktspektrums (abgesehen von $\lambda = 0$, $\varphi =$ Konstans) erwiesen.

4. Wir nehmen nun von $f(u)$, das wir periodisch mit der Periode 1 über alle u fortsetzen (so daß wir statt $f(\{u\})$ $f(u)$ schreiben können), folgendes an: Abgesehen von endlich vielen Punkten $x^{(1)}, \cdots, x^{(\nu)}$ (sowie den ihnen mod der Periode 1 kongruenten Punkten $x^{(1)}+k, \cdots, x^{(\nu)}+k$, $k = 0, \pm 1, \pm 2, \cdots$; $x^{(1)}, \cdots, x^{(\nu)}$ seien $\geqq 0, <1$) sei $f(u)$ überall stetig und stetig differentiierbar. In $x^{(1)}, \cdots, x^{(\nu)}$ möge es die bzw. Sprünge $\alpha^{(1)}, \cdots, \alpha^{(\nu)}$ machen. Die Summe dieser Sprünge, $\alpha = \alpha^{(1)} + \cdots + \alpha^{(\nu)}$, sei $\neq 0$.

Wir teilen das Intervall $0 \leqq u < 1$ in $(\nu+1)k$ Teile ein: I_l sei das Intervall $\dfrac{l-1}{(\nu+1)k} \leqq u < \dfrac{l}{(\nu+1)k}$, $l = 1, 2, \cdots, (\nu+1)k$. Die $\{\alpha^{(1)} - j\gamma\}, \cdots, \{\alpha^{(\nu)} - j\gamma\}$, $j = 0, 1, \cdots, k-1$ sind νk Zahlen, mindestens k unter den $(\nu+1)k$ Intervallen I_l sind also von ihnen frei, etwa: $I_{l^{(1)}}, \cdots, I_{l^{(k)}}$. In diesen sind also u, $\{u+\gamma\}, \cdots, \{u+(k-1)\gamma\}$ alle $\neq \alpha^{(1)}, \cdots, \alpha^{(\nu)}$, d. h. $f(u)+f(u+\gamma)+\cdots+f(u+(k-1)\gamma)$ stetig. Also ist in $I_{l^{(h)}}$, $h = 1, \cdots, k$,

$$f(u)+f(u+\gamma)+\cdots+f(u+(k-1)\gamma)$$
$$= C_h + \int_0^u \left(f'(x)+f'(x+\gamma)+\cdots+f'(x+(k-1)\gamma)\right) du.$$

Nun gilt für alle beschränkten, nach Riemann integrierbaren Funktionen $g(u)$ (mit der Periode 1)

$$\frac{g(u)+g(u+\gamma)+\cdots+g(u+(k-1)\gamma)}{k} \to \int_0^1 g(x)\,dx,$$

und zwar gleichmäßig in u.[70] Setzen wir für $g(x)$ $f'(x)$ ein und beachten

$$\int_0^1 f'(u)\,du + \alpha^{(1)} + \cdots + \alpha^{(\nu)} = 0, \quad \int_0^1 f'(u)\,du = -\alpha^{(1)} - \cdots - \alpha^{(\nu)} = -\alpha,$$

so sehen wir, daß

$$\frac{f'(u)+f'(u+\gamma)+\cdots+f'(u+(k-1)\gamma)}{k} \to -\alpha \quad \text{für} \quad k \to \infty$$

gleichmäßig in u gilt. Wegen $\alpha \neq 0$ hat also

$$\frac{f'(u)+f'(u+\gamma)+\cdots+f'(u+(k-1)\gamma)}{k}$$

[70] Weyl, a. a. O. Anm.[54]. Der dort geführte Beweis der Konvergenz ergibt auch die gleichmäßige Konvergenz.

ist, ist die Erfüllbarkeit der obigen Bedingungen für $\psi(u)$, zusammen mit der Endlichkeit von $\int_0^1 |\psi(u)|^2 du$, für den Eigenwert-Charakter von λ charakteristisch.

Für $\lambda = 0$ besagt dies $\psi(\{u+\gamma\}) = \psi(u)$. Nun bilden die $e^{2\pi i k u}$, $k = 0, \pm 1, \pm 2, \cdots$, ein vollständiges normiertes Orthogonalsystem, und sind Eigenfunktionen des unitären Operators $T\chi(u) = \chi(\{u+\gamma\})$, und zwar mit den bzw. Eigenwerten $e^{2\pi i k \gamma}$. Also gibt es keine weiteren Eigenfunktionen. Ist γ rational, so sind viele $e^{2\pi i k\gamma} = 1$, also viele $\psi(u) = e^{2\pi i k u}$ möglich, d. h. unsere Strömung nicht ergodisch. Ist es dagegen irrational, was im folgenden vorausgesetzt werde, so folgt aus $e^{2\pi i k\gamma} = 1$ $k = 0$, also muß $\psi(u)$ ein konstantes Vielfaches der einzigen Eigenfunktion, 1, sein — also $\psi(u) = c$, $\varphi(u, v) = c$, d. h. unsere Strömung ist ergodisch.

Sei nun $\lambda \neq 0$. Es ist $|\psi(\{u+\gamma\})| = |\psi(u)|$, also die obigen Schlüsse auf $|\psi(u)|$ anwendbar: es ist $|\psi(u)| = c$, wegen $\varphi \neq 0$, $\psi \neq 0$ ist $c > 0$. $\int_0^1 |\chi(\{u+\varepsilon\}) - \chi(u)|^2 du \to 0$ für $\varepsilon \to 0$ ist eine wohlbekannte Tatsache[68]. — Nach dem Satze von Kronecker[69] können wir, da γ irrational ist, eine Folge ganzer Zahlen $k_1 < k_2 < \cdots$ mit $\{k_n\gamma\} \to 0$ für $n \to \infty$ finden. Also ist dann

$$\int_0^1 |\psi(\{u+k_n\gamma\}) - \psi(u)|^2 du = \int_0^1 |\psi(\{u+\{k_n\gamma\}\}) - \psi(u)|^2 du \to 0.$$

Nun ist andererseits

$$\psi(\{u+k\gamma\}) = e^{i\lambda(f(u)+f(\{u+\gamma\})+\cdots+f(\{u+(k-1)\gamma\}))} \cdot \psi(u),$$

also

$$\begin{aligned}
&\int_0^1 |\psi(\{u+k\gamma\}) - \psi(u)|^2 du \\
&= \int_0^1 |(e^{i\lambda(f(u)+f(\{u+\gamma\})+\cdots+f(\{u+(k-1)\gamma\}))} - 1)\psi(u)|^2 du \\
&= c^2 \int_0^1 |e^{i\lambda(f(u)+f(\{u+\gamma\})+\cdots+f(\{u+(k-1)\gamma\}))} - 1|^2 du \\
&= 4c^2 \int_0^1 \sin^2\left(\frac{\lambda}{2}(f(u)+f(\{u+\gamma\})+\cdots+f(\{u+(k-1)\gamma\}))\right) du.
\end{aligned}$$

Wenn wir also

$$\operatorname{limes\,inf}_{k\to\infty} \left[\int_0^1 \sin^2\left(\frac{\lambda}{2}(f(u)+f(\{u+\gamma\})+\cdots+f(\{u+(k-1)\gamma\}))\right) du\right] > 0$$

[68] Es ist evident für alle $\chi(u) = e^{2\pi i k u}$, $k = 0, \pm 1, \pm 2, \cdots$, also gilt es auch für deren Linearaggregate und für die Häufungspunkte der Linearaggregate — also für alle $\chi(u)$.

[69] Vgl. z. B. bei Weyl, a. a. O. Anm. [54].

beweisen können, so ist die Unerfüllbarkeit der $\psi(u)$-Bedingung, d. h. die Nichtexistenz des Punktspektrums (abgesehen von $\lambda = 0$, $\varphi =$ Konstans) erwiesen.

4. Wir nehmen nun von $f(u)$, das wir periodisch mit der Periode 1 über alle u fortsetzen (so daß wir statt $f(\{u\})$ $f(u)$ schreiben können), folgendes an: Abgesehen von endlich vielen Punkten $x^{(1)}, \cdots, x^{(\nu)}$ (sowie den ihnen mod der Periode 1 kongruenten Punkten $x^{(1)}+k, \cdots, x^{(\nu)}+k$, $k = 0, \pm 1, \pm 2, \cdots$; $x^{(1)}, \cdots, x^{(\nu)}$ seien $\geqq 0, <1$) sei $f(u)$ überall stetig und stetig differentiierbar. In $x^{(1)}, \cdots, x^{(\nu)}$ möge es die bzw. Sprünge $\alpha^{(1)}, \cdots, \alpha^{(\nu)}$ machen. Die Summe dieser Sprünge, $\alpha = \alpha^{(1)} + \cdots + \alpha^{(\nu)}$, sei $\neq 0$.

Wir teilen das Intervall $0 \leqq u < 1$ in $(\nu+1)k$ Teile ein: I_l sei das Intervall $\dfrac{l-1}{(\nu+1)k} \leqq u < \dfrac{l}{(\nu+1)k}$, $l = 1, 2, \cdots, (\nu+1)k$. Die $\{\alpha^{(1)} - j\gamma\}, \cdots, \{\alpha^{(\nu)} - j\gamma\}$, $j = 0, 1, \cdots, k-1$ sind νk Zahlen, mindestens k unter den $(\nu+1)k$ Intervallen I_l sind also von ihnen frei, etwa: $I_{l^{(1)}}, \cdots, I_{l^{(k)}}$. In diesen sind also u, $\{u+\gamma\}, \cdots, \{u+(k-1)\gamma\}$ alle $\neq \alpha^{(1)}, \cdots, \alpha^{(\nu)}$, d. h. $f(u)+f(u+\gamma)+\cdots+f(u+(k-1)\gamma)$ stetig. Also ist in $I_{l^{(h)}}$, $h = 1, \cdots, k$,

$$f(u)+f(u+\gamma)+\cdots+f(u+(k-1)\gamma)$$
$$= C_h + \int_0^u \left(f'(x)+f'(x+\gamma)+\cdots+f'(x+(k-1)\gamma)\right) du.$$

Nun gilt für alle beschränkten, nach Riemann integrierbaren Funktionen $g(u)$ (mit der Periode 1)

$$\frac{g(u)+g(u+\gamma)+\cdots+g(u+(k-1)\gamma)}{k} \to \int_0^1 g(x)\, dx,$$

und zwar gleichmäßig in u.[70] Setzen wir für $g(x)$ $f'(x)$ ein und beachten

$$\int_0^1 f'(u)\, du + \alpha^{(1)} + \cdots + \alpha^{(\nu)} = 0, \quad \int_0^1 f'(u)\, du = -\alpha^{(1)} - \cdots - \alpha^{(\nu)} = -\alpha,$$

so sehen wir, daß

$$\frac{f'(u)+f'(u+\gamma)+\cdots+f'(u+(k-1)\gamma)}{k} \to -\alpha \quad \text{für} \quad k \to \infty$$

gleichmäßig in u gilt. Wegen $\alpha \neq 0$ hat also

$$\frac{f'(u)+f'(u+\gamma)+\cdots+f'(u+(k-1)\gamma)}{k}$$

[70] Weyl, a. a. O. Anm.[54]. Der dort geführte Beweis der Konvergenz ergibt auch die gleichmäßige Konvergenz.

bei hinreichend großem k für alle u dasselbe Vorzeichen und einen Absolutwert $\geqq \frac{1}{2}|\alpha|$, $\leqq \frac{3}{2}|\alpha|$.

Also ist in $I_l^{(h)}$

$$f(u)+f(u+\gamma)+\cdots+f(u+(k-1)\gamma) = C_h \pm \frac{|\alpha|k}{2}\int_0^u \sigma_h(x)\,dx,$$

wobei stets $1 \leqq \sigma_h(x) \leqq 3$ ist. Die Gesamtvariation dieses Ausdrucks in $I_l^{(h)}$, dessen Länge $\frac{1}{(\nu+1)k}$ ist, ist $\leqq \frac{1}{(\nu+1)k}\cdot\frac{|\alpha|k}{2}\cdot 3 = \frac{3|\alpha|}{\nu+1}$, also $< \left|\frac{\pi}{\lambda}\right|$, wenn ν groß genug ist. Dies darf aber vorausgesetzt werden, denn wir können ν durch Hinzufügen fiktiver Unstetigkeitspunkte (mit Sprüngen $= 0$) beliebig erhöhen. Die Gesamtvariation ist also $<$ als die halbe Distanz von zwei benachbarten $\frac{2\pi n}{\lambda}$, $n = 0, \pm 1, \pm 2, \cdots$. Ferner variiert der obige Ausdruck in ganz $I_l^{(h)}$ monoton. Mindestens in einer Hälfte von $I_l^{(h)}$ variiert er also (bei wachsendem u) vom nächsten $\frac{2\pi n}{\lambda}$ weg. Da er ferner im Viertel von $I_l^{(h)}$ um $\geqq \frac{1}{4}\cdot\frac{1}{(\nu+1)k}\cdot\frac{|\alpha|k}{2} = \frac{|\alpha|}{8(\nu+1)}$ wächst, ist er mindestens in einem Viertel von $I_l^{(h)}$ um $\geqq \frac{|\alpha|}{8(\nu+1)}$ vom nächsten $\frac{2\pi n}{\lambda}$ entfernt. Sein $\frac{\lambda}{2}$-faches, das Argument im $\sin^2(\cdots)$ des $\underset{k\to\infty}{\text{limes inf}}\left[\int_0^1 \sin^2(\cdots)du\right]$ am Ende von 3., ist also dort um $\geqq \frac{|\alpha|\cdot|\lambda|}{16(\nu+1)}$ vom nächsten πn entfernt. Da aber $\sin^2 y$ stets $\geqq \left(\frac{\text{2. Distanz des nächsten } \pi n \text{ von } y}{\pi}\right)^2$ ist, ist der $\sin^2(\cdots)$ selbst $\geqq \frac{\alpha^2\lambda^2}{64\pi^2(\nu+1)^2}$.

Im Integral $\int_0^1 \sin^2(\cdots)\,du$ am Ende von 3. sind also, falls k groß genug ist, k Intervalle (die genannten Viertel der $I_l^{(h)}$, $h = 1, \cdots, k$) vorhanden, jedes von der Länge $\frac{1}{4(\nu+1)k}$, in deren jedem der Integrand $\geqq \frac{\alpha^2\lambda^2}{64\pi^2(\nu+1)^2}$ ist. Da der Integrand stets $\geqq 0$ ist, ist das ganze Integral jedenfalls

$$\geqq k\cdot\frac{1}{4(\nu+1)k}\cdot\frac{\alpha^2\lambda^2}{64\pi^2(\nu+1)^2} = \frac{\alpha^2\lambda^2}{256\pi^2(\nu+1)^3} = \delta > 0,$$

also

$$\underset{k\to\infty}{\text{limes inf}}\left[\int_0^1 \sin^2(\cdots)\,du\right] \geqq \delta > 0.$$

Damit ist das am Ende von 3. Gesagte bewiesen.

5. Wir haben also gezeigt:

SATZ 6. *Sei S_t die in 1. mit Hilfe der Funktion $f(u)$ (die periodisch mit der Periode 1 fortgesetzt zu denken ist) und der Zahl γ definierte Strömung. $f(u)$ sei überall stetig und stetig differentiierbar, abgesehen von endlich vielen Stellen in $0 \leqq u < 1$ (sowie den diesen mod der Periode 1 kongruenten Stellen), wo Unstetigkeiten in Form von Sprüngen (d. i. $f(u+0)$, $f(u-0)$ existieren, aber $f(u+0)-f(u-0) \neq 0$) zulässig sind. Die Derivierte sei beschränkt und nach Riemann integrierbar. Ferner sei $0 \leqq \gamma < 1$.*

Dann ist für den (eigentlich-) ergodischen Charakter die Irrationalität von γ notwendig und hinreichend; für die Nichtexistenz eines Punktspektrums (außer $\lambda = 0$, $\varphi =$ Konstans) das Nichtverschwinden der Summe der Sprünge von $f(u)$ (in $0 \leqq u < 1$) hinreichend.

Das reine Streckenspektrum ist also hier der allgemeine Fall.

Wie weit unsere Bedingung notwendig ist, ist schwerer zu entscheiden. Wenn $f(u)$ überall stetig ist, oder die Summe seiner Sprünge Null ist, kann ein Punktspektrum auftreten. Z. B.:

Sei $f(u)$ ein trigonometrisches Polynom:

$$f(u) = a_0 + \sum_1^n{}^m (a_m \cos 2\pi m u + b_m \sin 2\pi m u).$$

(Wegen $A \leqq f(u) \leqq B$ ist $A \leqq a_0 \leqq B$). Damit $\psi(\{u+\gamma\}) = e^{i\lambda f(u)}\,\psi(u)$ lösbar sei, genügt die Lösbarkeit von $\chi(\{u+\gamma\}) - \chi(u) = f(u) \cdots \bmod \frac{2\pi}{\lambda}$, da wir $\psi(u) = e^{i\lambda\chi(u)}$ setzen können. Da aber

$$\chi_1(\{u+\gamma\}) - \chi_1(u) = \sum_1^n{}^m (a_m \cos 2\pi m u + b_m \sin 2\pi m u)$$

lösbar ist[71], genügt es $\chi(\{u+\gamma\}) - \chi(u) = a_0 \cdots \bmod \frac{2\pi}{\lambda}$ zu lösen. Setzen wir $\chi(u) = \alpha u$ an, so ist

$$\chi(\{u+\gamma\}) - \chi(u) = \alpha\gamma \text{ bzw. } = \alpha(\gamma - 1),$$

also muß dann

$$\alpha\gamma - a_0 = \frac{2\pi k}{\lambda}, \qquad \alpha\gamma - \alpha - a_0 = \frac{2\pi l}{\lambda} \quad (k, l = 0, \pm 1, \pm 2, \cdots)$$

sein, also

---

[71] Es genügt $\chi_1(\{u+\gamma\}) - \chi_1(u) = \cos 2\pi m u$ und $= \sin 2\pi m u$ zu lösen, d. h. $= \Re e^{2\pi i m u}$ und $= \Im e^{2\pi i m u}$. Dies leistet

$$\chi_1(u) = \Re \text{ bzw. } \Im \frac{e^{2\pi i m u}}{e^{2\pi i m \gamma} - 1}.$$

$$\alpha = \frac{2\pi(k-l)}{\lambda}, \text{ und } a_0 = \alpha\gamma - \frac{2\pi k}{\lambda} = \frac{2\pi\gamma(k-l)-2\pi k}{\lambda},$$

$$\lambda = \frac{2\pi\gamma}{a_0}(k-l) - \frac{2\pi}{a_0}k.$$

Wenn wir also statt $k-l, -k$ k, l schreiben, sehen wir: die

$$\lambda = \frac{2\pi\gamma}{a_0}k + \frac{2\pi}{a_0}l \quad (k, l = 0, \pm 1, \pm 2, \cdots)$$

gehören alle zum Punktspektrum.

Oder es sei $f(u)$ $\begin{cases} = 1 \text{ für } 0 \leqq u < \frac{1}{2} \\ = 2 \text{ für } \frac{1}{2} \leqq u < 1 \end{cases}$. (Die Sprünge sind: -1 bei $u = 0$, 1 bei $u = \frac{1}{2}$, ihre Summe ist 0.) Wenn $\lambda = 2\pi k$ $(k = 0, \pm 1, \pm 2, \cdots)$ ist, ist $e^{i\lambda f(u)} = 1$, also $\psi(u) =$ Konstans Lösung.

Andererseits gibt es stetige $f(u)$, für die ein reines Streckenspektrum vorliegt, z. B.

$$f(u) = a_0 + a_1 \sum_1^\infty{}^n \frac{\cos(2^{(2^n)}\pi u)}{2^{(2^n)}} \qquad \left(a_0 > |a_1| \sum_1^\infty{}^n \frac{1}{2^{(2^n)}}\right),$$

falls in der dyadischen Entwicklung von γ für unendlich viele m die $2^{(2^m)}$-te Ziffer eine 1 ist, und alle weiteren Ziffern, bis zur $2^{2^{(m+1)}}$-ten (ausschl.), Nullen. Man beweist dies durch explizite Ausrechnung von $f(u)+f(u+\gamma)+\cdots+f(u+(k-1)\gamma)$ für $k = 2^{(2^m)}$, und Abschätzung des Einflusses der einzelnen Glieder von $\sum_1^\infty{}^n$, wobei dasjenige mit $n = m$ dominiert — dann zeigt sich, daß der $\limes_{m\to\infty}\inf\left[\int_0^1 \sin^2(\cdots)\,du\right]$ vom Ende von 3. >0 ist, obwohl $\{2^{(2^m)}\gamma\}$ nach Annahme beliebig kleine Werte annimmt. Wir übergehen die Einzelheiten dieser Rechnung. Vermutlich liegt bei den meisten stetigen $f(u)$ ein reines Streckenspektrum vor.

Unsere Strömung S_t läßt sich noch auf anschaulichere Formen bringen (durch isomorphe Abbildungen), doch soll dies bei einer anderen Gelegenheit erörtert werden.

ON ONE-PARAMETER UNITARY GROUPS IN HILBERT SPACE.[1]

By M. H. Stone.

In this note, I give the detailed proof of a result (Theorem B below) which I announced in the Proceedings of the National Academy of Sciences, 16 (1930), pp. 173–4 and which I originally intended to publish in my forthcoming book, "Linear Transformations in Hilbert Space and Their Applications to Analysis" (Colloquium Publications of the American Mathematical Society). The chapter on group theory was completed in May, 1930, but was later excluded from the book. Recent applications of this theorem make further delay in publication highly undesirable. J. v. Neumann, for example, has found it necessary to generalize the statement of the theorem and has given a new proof, which appears in this number of the Annals.[2] My own method seems to me to be a "correct" one from the point of view of the theory of groups and their linear representations; and in any case, as I had observed in my original manuscript, it yields without any modification the generalized form of the theorem required by v. Neumann (Theorem C below). The numbered definitions, theorems, and lemmas to which I refer in these pages, will be found in my book.

THEOREM A. *If H is a self-adjoint transformation in abstract Hilbert space $\mathfrak{H}$, then $U(\tau) \equiv e^{i\tau H}$, $-\infty < \tau < +\infty$, is a family of unitary transformations with the group property*

$$(\alpha) \qquad U(\sigma+\tau) \equiv U(\sigma)\,U(\tau)$$

and the continuity property

$$(\beta) \qquad U(\sigma) \to U(\tau) \quad \textit{when} \quad \sigma \to \tau \text{ (Def. 2.3).}$$

The existence of $U(\tau) \equiv e^{i\tau H}$ is established in Th. 6.1, its unitary character follows from Th. 6.6. The group property (α) is a consequence of Th. 6.1, (6), (7). The continuity property (β) follows from the relation

$$\|(U(\sigma) - U(\tau))f\|^2 = \int_{-\infty}^{+\infty} |e^{i\sigma\lambda} - e^{i\tau\lambda}|^2 \, d\|E(\lambda)f\|^2,$$

where $E(\lambda)$ is the resolution of the identity corresponding to H, as we have pointed out in Th. 6.1 (5) and Th. 6.2.

[1] Received March 19, 1932.

[2] Vide supra, pp. 567–573.

THEOREM B. *If* $U(\tau)$, $-\infty<\tau<+\infty$, *is a family of unitary transformations with the group property* (α) *and the continuity property* (β), *then there exists a unique self-adjoint transformation* H *such that* $U(\tau)\equiv e^{i\tau H}$.

It is evident that the continuity property (β) implies the continuity of $(U(\tau)f, g)$ as a function of τ for arbitrary f and g in $\mathfrak{H}$, by virtue of Def. 2.3 and Th. 1.3. Hence we may generalize Theorem B as follows:

THEOREM C. *If* $U(\tau)$, $-\infty<\tau<+\infty$, *is a family of unitary transformations with the group property* (α) *and the measurability property*

(γ) $(U(\tau)f, g)$ *is a measurable function of* τ *for arbitrary* f *and* g *in* $\mathfrak{H}$, *then there exists a unique self-adjoint transformation* H *such that* $U(\tau)\equiv e^{i\tau H}$. *Theorem* A *shows, therefore, that the group property* (α) *and the measurability property* (γ) *together imply the continuity property* (β).

We now prove Theorem C. First, by the Fourier-Plancherel Theorem (Th. 3.10), the reciprocal formulas

$$\psi(\tau; l) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{1}{\lambda-l}e^{-i\tau\lambda}\,d\lambda, \tag{1a}$$

$$\frac{1}{\lambda-l} = \int_{-\infty}^{+\infty}\psi(\tau; l)\,e^{i\lambda\tau}\,d\tau \tag{1b}$$

are valid when $\Im(l)\neq 0$, and lead to the relation

$$(l-m)\int_{-\infty}^{+\infty}\psi(\tau; l)\,\psi(\sigma-\tau; m)\,d\tau = \psi(\sigma; l)-\psi(\sigma; m) \tag{2}$$

when the unitary character of the Fourier-Plancherel transformation is coupled with equation (1a) and the equation obtained from (1a) by the substitution of $\sigma-\tau$ for τ. We find

$$\begin{aligned}&\psi(\tau; l)=0, && \tau>0,\ \Im(l)>0; && \psi(\tau; l)=ie^{-il\tau}, && \tau<0,\ \Im(l)>0,\\ &\psi(\tau; l)=-ie^{-il\tau}, && \tau>0,\ \Im(l)<0; && \psi(\tau; l)=0, && \tau<0,\ \Im(l)<0;\end{aligned} \tag{3}$$

since (1b) is obviously valid with this choice of ψ. We find also that

$$\overline{\psi(\tau; l)} = \psi(-\tau; \bar{l}). \tag{4}$$

A more complete discussion of some of these relations is given in the proof of Th. 10.9. Secondly, we define a bounded linear transformation X_l with domain $\mathfrak{H}$ by the equation

$$(X_l f, g) = \int_{-\infty}^{+\infty}\psi(\tau; l)\,(U(\tau)f, g)\,d\tau, \qquad \Im(l)\neq 0.$$

The integral is to be interpreted as a Lebesgue integral in accordance with the measurability property (γ). By the unitary character of $U(\tau)$

we have $|(U(\tau)f, g)| \leqq |U(\tau)f|\,|g| = |f|\,|g|$. By combining this inequality with (3), we find that the integral is absolutely convergent and does not exceed $|f|\,|g|/|\Im(l)|$ in absolute value. We can therefore apply a theorem of Fréchet (Ths. 2.27–2.28) to establish the existence of the desired transformation X_l. Thirdly, we obtain essential properties of X_l corresponding to (2), (3), and (4). The first is

(2 bis) $$(l-m)\,X_l\,X_m \equiv X_l - X_m, \quad \Im(l) \neq 0, \quad \Im(m) \neq 0.$$

From the definition of X_l and the group property (α), we have

$$\begin{aligned}(X_l\,X_m f, g) = (X_m f, X_l^* g) &= \int_{-\infty}^{+\infty} \psi(\varrho; m)\,(U(\varrho)f, X_l^* g)\,d\varrho \\ &= \int_{-\infty}^{+\infty} \psi(\varrho; m)\,(X_l\,U(\varrho)f, g)\,d\varrho \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \psi(\tau; l)\,\psi(\varrho; m)\,(U(\tau)\,U(\varrho)f, g)\,d\tau\,d\varrho \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \psi(\tau; l)\,\psi(\varrho; m)\,(U(\varrho+\tau)f, g)\,d\tau\,d\varrho.\end{aligned}$$

This integral is an absolutely convergent Lebesgue integral by virtue of the measurability property (γ),[3] the relations (3), and the inequality $|(U(\varrho+\tau)f, g)| \leqq |f|\,|g|$. Hence we can introduce new variables $\sigma = \varrho + \tau$ and $\tau = \tau$ and can integrate in any convenient order. If we perform the integration with respect to the new variable τ first, we obtain with the aid of (2) the result

$$(l-m)\,(X_l\,X_m f, g) = \int_{-\infty}^{+\infty} (\psi(\sigma; l) - \psi(\sigma; m))\,(U(\sigma)f, g)\,d\sigma = ((X_l - X_m)f, g),$$

which evidently implies (2 bis). The second property is

(3 bis) $$X_m f = 0 \quad \text{implies} \quad f = 0.$$

From (2 bis) we see that $X_m f = 0$ implies $X_l f = 0$ for all not-real l and hence implies that $\int_{-\infty}^{+\infty} \psi(\tau; l)\,(U(\tau)f, g)\,d\tau$ vanishes for all not-real l and all g in $\mathfrak{H}$. By making use of the explicit form of ψ given in (3), we conclude[4] that $(U(\tau)f, g) = 0$ for almost all τ and all g in $\mathfrak{H}$. Hence

[3] It is to be noted that if $f(\tau)$ is a measurable function of τ, $-\infty < \tau < +\infty$, then $F(\varrho, \tau) = f(\varrho + \tau)$ is a measurable function of (ϱ, τ).

[4] We have to show that the equation $\int_0^{+\infty} e^{il\tau} F(\tau)\,d\tau = 0$, holding for all l such that $\Im(l) > 0$, implies $F(\tau) = 0$ almost everywhere, $\tau > 0$, whenever $F(\tau)$ is bounded and measurable. If we put $l = \mu + i\nu$, $\nu > 0$, and hold ν fast, we can obtain this result by a simple application of the Fourier-Plancherel Theorem.

we see that $U(\tau)f = 0$ for almost all τ and that $f = 0$ by virtue of the unitary character of $U(\tau)$. The third property is

(4 bis) $$X_l^* \equiv X_{\bar{l}}, \qquad \Im(l) \neq 0.$$

The group property (α) implies that $U(0) \equiv I$, $U(-\tau) \equiv U^{-1}(\tau) \equiv U^*(\tau)$. Hence we obtain from (4) the relation

$$\begin{aligned}(X_l^* f, g) = \overline{(X_l g, f)} &= \int_{-\infty}^{+\infty} \overline{\psi(\tau; l)}\, \overline{(U(\tau) g, f)}\, d\tau \\ &= \int_{-\infty}^{+\infty} \psi(-\tau; \bar{l})\, (U^*(\tau) f, g)\, d\tau \\ &= \int_{-\infty}^{+\infty} \psi(\tau; \bar{l})\, (U(\tau) f, g)\, d\tau \\ &= (X_{\bar{l}} f, g),\end{aligned}$$

which holds for all f and g in $\mathfrak{H}$ and therefore implies (4 bis). In the fourth place, we observe that, in accordance with Th. 4.19, the three properties of X_l which we have just proved enable us to assert the existence of a unique self-adjoint transformation H such that $X_l \equiv (H-l)^{-1}$ for all not-real l. Finally we must show that the transformation $V(\tau) \equiv e^{i\tau H}$ coincides with $U(\tau)$. According to Theorem A, the family $V(\tau)$ satisfies the hypotheses of Theorems B and C. Hence we can introduce the transformation Y_l defined by the relation

$$(Y_l f, g) = \int_{-\infty}^{+\infty} \psi(\tau; l)\, (V(\tau) f, g)\, d\tau.$$

By Th. 6.1, we have

$$\begin{aligned}(Y_l f, g) &= \lim_{a \to +\infty} \int_{-a}^{+a} \psi(\tau; l) \left(\int_{-\infty}^{+\infty} e^{i\lambda\tau}\, d(E(\lambda) f, g) \right) d\tau \\ &= \lim_{a \to +\infty} \int_{-\infty}^{+\infty} \left(\int_{-a}^{+a} e^{i\lambda\tau}\, \psi(\tau; l)\, d\tau \right) d(E(\lambda) f, g) \\ &= \lim_{a \to +\infty} \int_{-\infty}^{+\infty} \frac{1 - e^{\pm ia(\lambda - l)}}{\lambda - l}\, d(E(\lambda) f, g) \\ &= \int_{-\infty}^{+\infty} \lim_{a \to +\infty} \frac{1 - e^{\pm ia(\lambda - l)}}{\lambda - l}\, d(E(\lambda) f, g) \\ &= \int_{-\infty}^{+\infty} \frac{1}{\lambda - l}\, d(E(\lambda) f, g),\end{aligned}$$

where $E(\lambda)$ is the resolution of the identity associated with H, and the $+$ sign or the $-$ sign is taken according as $\Im(l) < 0$ or $\Im(l) > 0$. Now Ths. 5.7 and 6.1 show that $Y_l \equiv (H-l)^{-1} \equiv X_l$. Hence we have

$$\int_{-\infty}^{+\infty} \psi(\tau; l)\,((U(\tau)f, g) - (V(\tau)f, g))\,d\tau = 0$$

and can conclude that $U(\tau) \equiv V(\tau)$ for almost all τ. If τ is fixed and ϱ varies in the set specified by the identity $U(\varrho) \equiv V(\varrho)$, then the difference $\sigma = \tau - \varrho$ varies over the entire range $-\infty < \sigma < +\infty$ with the possible exception of a null set. Hence σ assumes at least one value such that $U(\sigma) \equiv V(\sigma)$. We see therefore that

$$U(\tau) \equiv U(\varrho+\sigma) \equiv U(\varrho)U(\sigma) \equiv V(\varrho)V(\sigma) \equiv V(\varrho+\sigma) \equiv V(\tau)$$

for arbitrary τ and suitably chosen ϱ and σ, by virtue of the group property (α). This completes the proof.

THEOREM D. *If H is a self-adjoint transformation and $U(\tau) \equiv e^{i\tau H}$, then the relation $\dfrac{U(\tau)-U(0)}{\tau} f \to f^*$ is valid in $\mathfrak{H}$ if and only if f is in the domain of H and $f^* = iHf$.*

If f is in the domain of H, the integral $\int_{-\infty}^{+\infty} \lambda^2 d|E(\lambda)f|^2$ exists by Th. 5.9. From Th. 6.1 (5) and Lemma 6.1 (6) we see therefore that

$$\lim_{\tau\to 0} \left\| \left(H - \frac{e^{i\tau H}-1}{i\tau}\right) f \right\|^2 = \lim_{\tau\to 0} \int_{-\infty}^{+\infty} \lambda^2 \left| 1 - \frac{e^{i\tau\lambda}-1}{i\tau\lambda} \right|^2 d\|E(\lambda)f\|^2$$

$$= \int_{-\infty}^{+\infty} \lim_{\tau\to 0} \lambda^2 \left| 1 - \frac{e^{i\tau\lambda}-1}{i\tau\lambda} \right|^2 d\|E(\lambda)f\|^2 = 0.$$

This result means that $\dfrac{U(\tau)-U(0)}{\tau} f \to f^* = iHf$ whenever f is in the domain of H. On the other hand, we can define a transformation T by the relation $iTf = f^*$, valid whenever $\dfrac{U(\tau)-U(0)}{\tau} f$ converges in $\mathfrak{H}$ to an element f^*. Then T is an extension of H with the property that

$$(Tf, g) = \lim_{\tau\to 0} \left(\frac{U(\tau)-U(0)}{i\tau} f, g \right)$$

$$= \lim_{\tau\to 0} \left(f, \frac{U(-\tau)-U(0)}{-i\tau} g \right) = (f, Tg)$$

whenever f and g are elements in its domain. Hence T is a symmetric extension of H. Since H is self-adjoint, we must have $T \equiv H$ in accordance with Th. 2.13. For a modification of this theorem, we refer to v. Neumann's paper in this number of the Annals.

THEOREM E. *A bounded linear transformation T is permutable (Def. 8.1) with the self-adjoint transformation H if and only if it is permutable with every transformation of the family $e^{i\tau H}$, $-\infty < \tau < +\infty$.*

From the integral relation connecting $(H - l)^{-1}$ with $e^{i\tau H}$, it is evident that T is permutable with every transformation $e^{i\tau H}$ if and only if it is permutable with $(H - l)^{-1}$ for all not-real l. As we proved in Th. 8.1, T is permutable with $(H - l)^{-1}$ for all not-real l (or, for a single such l) if and only if it is permutable with H.

Because of lack of space, I do not go into certain other generalizations of the theorems proved here. I point out that the results obtained lead easily to a characterization of all n-parameter abelian unitary groups $U(\tau_1, \cdots, \tau_n) \equiv e^{i\tau_1 H_1} \cdots e^{i\tau_n H_n}$, where $H_1, \cdots, H_n$ are permutable self-adjoint transformations (cf. Def. 8.2). A less obvious generalization deals with the family of isometric transformations $U(\tau)$, $0 \leqq \tau < +\infty$, where $U(\sigma + \tau) \equiv U(\sigma)\, U(\tau)$ for $\sigma \geqq 0$, $\tau \geqq 0$. I hope to state this problem in precise form and to study its solution on some other occasion.

YALE UNIVERSITY.

PAIRS OF CURVES IN AN S_n.*

By G. A. Parkinson.

Introduction. In 1845 Saint Venant[1] proposed the question whether upon the surface generated by the principal normal of a curve, a second curve can exist which has for its principal normal the principal normal of the given curve.

This question was answered by Bertrand,[2] in 1850 in a paper in which he showed that a necessary and sufficient condition for the existence of such a second curve is that a linear relationship with constant coefficients shall exist between the first and second curvatures of the given original curve.

Since the time of Bertrand's paper, pairs of curves of this kind have been called *Conjugate Bertrand Curves,* or more commonly, *Bertrand Curves.*

The purpose of the present paper is to investigate in a general flat space S_n, of n dimensions, the properties of pairs of curves which have a common pth normal, where p may have any value from 1 to $n-1$; pairs of curves for which the principal normal of one lies in the osculating plane of the other and pairs of curves for which the principal normal of one lies in the plane of the principal normal and binormal of the other.

1. **The Frenet formulas for curves in an S_n.**[3] Let

$$x = x(s) \tag{1.1}$$

be the equations of a curve C in an S_n, where s is the arc length of C measured from an arbitrary point and where the x are n single valued analytic functions of s.

A set of n linearly independent vectors ξ[4] are defined along C subject to the conditions that ξ_1 shall be the unit tangent to the curve and such that $\xi_{\alpha+1} = d\xi_\alpha/ds$.

* Received March 26, and July 14, 1931. — This paper was prepared at the suggestion of Professor James H. Taylor, whose kind assistance has been appreciated by the author.

[1] Saint Venant, Mémoire sur les lignes courbes non planes. Journal de l'Ecole polyt., cah. 30, 1845, pp. 1–76.

[2] Bertrand, Mémoire sur la théorie des courbes à double courbure. Comptes Rendus, vol. 36, 1850, and Journal de Mathématiques, series I, vol. 15, pp. 332–350.

[3] Cf. Eisenhart, Riemannian Geometry, pp. 103 et. seq.

[4] In this paper a Greek letter represents a set of n vectors. The pth vector of the set is indicated by attaching the subscript p to the letter representing the set.

We denote by b_p the determinant $|b^\alpha_\beta|$ for $\alpha, \beta = 1, \cdots, p$ where,

$$b^\alpha_\beta = \xi_\alpha \cdot \xi_\beta. \tag{1.2}$$

A set of unitary orthogonal vectors μ are now defined along C by the equations,

$$\mu_p = \sqrt{\frac{e_p\, b_p}{b_{p-1}}}\, \xi_\alpha \cdot B^\alpha_p \qquad (\alpha \text{ is summed from 1 to } p). \tag{1.3}$$

In this equation $e_p = 1$ or -1 and is chosen so that the expression under the radical is positive. B^α_p is the cofactor of b^α_p in b_p divided by b_p. We take b_0 equal to 1 so that $\mu_1 = \xi_1$. Hence μ_1 is the unit tangent to C and we call μ_2 the principal normal, μ_3 the binormal, and in general μ_{p+1} the pth normal of C. This set of vectors is taken as the base of the coördinate system along C.

From the definition of the vectors μ it follows that,

$$\frac{d\mu_p}{ds} = -\frac{e_{p-1}}{\varrho_{p-1}}\mu_{p-1} + \frac{e_{p+1}}{\varrho_p}\mu_{p+1}. \tag{1.4}$$

These are the Frenet Formulas for C. The curvatures of C are defined by the equations,

$$\begin{cases} \dfrac{1}{\varrho_p} = \dfrac{1}{b_p}\sqrt{e_p\, e_{p+1}\, b_{p-1}\, b_{p+1}} & (p = 1, \cdots, n-1). \\ \dfrac{1}{\varrho_0} = \dfrac{1}{\varrho_n} = 0. \end{cases} \tag{1.5}$$

2. **Curves in a sub-space S_r of S_n $(r < n)$.** If a curve C lies in an S_r, there exist constants a_{ij} and b_i such that,

$$a_{ij}\, x_j = b_i \qquad \begin{pmatrix} i = 1, \cdots, n-r \\ j = 1, \cdots, n \end{pmatrix}. \tag{2.1}$$

By differentiation we obtain,

$$a_{ij}\frac{d^p x_j}{ds^p} = 0 \qquad (j = 1, \cdots, n), \tag{2.2}$$

where the matrix $\left\| \dfrac{d^p x_j}{ds^p} \right\|$ is at most of rank r.

The order of the x's may be chosen so that their derivatives may be expressed linearly in terms of the derivatives,

$$\frac{dx}{ds}, \quad \frac{d^2x}{ds^2}, \quad \cdots, \quad \frac{d^r x}{dx^r}. \tag{2.3}$$

Conversely if A's exist (not necessarily constant) such that,

$$\frac{d^p x}{ds^p} = A_i \frac{d^i x}{ds^i} \qquad \begin{pmatrix} i = 1, \cdots, r \\ p = r+1, \cdots, n \end{pmatrix} \tag{2.4}$$

the curve C lies in an S_r. For,

$$x^i = x^i(s_0) + \left(\frac{dx_i}{ds}\right)_0 (s-s_0) + \frac{1}{2}\left(\frac{d^2 x_i}{ds^2}\right)_0 (s-s_0)^2 + \cdots. \tag{2.5}$$

Hence,

$$x^i - x^i(s_0) = B_j \left(\frac{d^j x_i}{ds^j}\right)_0 \qquad (j = 1, \cdots, r), \tag{2.6}$$

where the B's are a power series in $(s-s_0)$. From this we see that a line joining a variable point of C to a fixed point (s_0) always lies in the S_r determined by the r fixed directions $\left(\frac{d^j x}{ds^j}\right)_0$, $(j = 1, \cdots, r)$ at the point s_0.

We now take,

$$\xi_\alpha = \frac{d^\alpha x}{ds^\alpha} \qquad (\alpha = 1, \cdots, r), \tag{2.7}$$

and define μ_p $(p = 1, \cdots, r)$ by equations analogous to (1.3). Let μ_p $(p = r+1, \cdots, n)$ be any set of directions which form with $\mu_1, \cdots, \mu_r$ an orthogonal n-nuple. The curvatures $1/\varrho_1, \cdots, 1/\varrho_r$ are defined by equations analogous to (1.5), and $1/\varrho_{r+1}, \cdots, 1/\varrho_n$ are defined as zero.

In this case equations (1.4) hold since here we have only restricted the choice of the ξ_α.

It follows from (1.4) that if $1/\varrho_p = 0$, the derivatives $d^k x/ds^k$ $(k = 1, \cdots, p+1)$ are linearly dependent and we have: *If the p'th curvature of a curve is zero the curve lies in at most an S_p.*

3. **Pairs of curves which have a common pth normal.** In this portion of the paper it is assumed that $a \neq 0$ and is positive in the direction of the center of pth curvature of C, and that neither $1/\varrho_p$ nor $ds/d\bar{s}$ is either zero or infinite.

If C and $\bar{C}$ have a common pth normal, the equations of C are,

$$\bar{x}(\bar{s}) = x(s) + a\mu_{p+1}. \tag{3.1}$$

In these equations, a is a scalar function of s, and s and $\bar{s}$ are the arc lengths of C and $\bar{C}$ respectively.

Differentiating both sides of this equation with respect to $\bar{s}$,

$$\frac{d\bar{x}}{d\bar{s}} = \left[\mu_1 + a\frac{d\mu_{p+1}}{ds} + a'\mu_{p+1}\right]\frac{ds}{d\bar{s}}.$$

Making use of the Frenet Formulas, this reduces to,

$$\bar{\mu}_1 = \left[\mu_1 + a\left(-\frac{e_p}{\varrho_p}\mu_p + \frac{e_{p+2}}{\varrho_{p+1}}\mu_{p+2}\right) + a'\mu_{p+1}\right]\frac{ds}{d\bar{s}}. \tag{3.2}$$

But $\bar{\mu}_1$ is orthogonal to $\bar{\mu}_{p+1}$ and $\bar{\mu}_{p+1} \equiv \mu_{p+1}$, hence $a' = 0$ and we have,

$$a = \text{constant}. \tag{3.3}$$

If two curves C and $\bar{C}$ have a common pth normal, they cut off on this normal a segment of constant length.

By (3.3), equation (3.2) reduces to,

$$\bar{\mu}_1 = \left[\mu_1 - \frac{e_p a}{\varrho_p}\mu_p + \frac{e_{p+2}a}{\varrho_{p+1}}\mu_{p+2}\right]\frac{ds}{d\bar{s}}. \tag{3.4}$$

Taking the scalar product of each side of this equation by itself and solving for $ds/d\bar{s}$ we have,

$$\frac{ds}{d\bar{s}} = \left\{\bar{e}_1\left[e_1 + \frac{e_p a^2}{\varrho_p^2} + \frac{e_{p+2}a^2}{\varrho_{p+1}^2}\right]\right\}^{-1/2} \equiv \Omega^{-1/2} > 0. \tag{3.5}$$

If we take the scalar product of each side of (3.4) by μ_1 and call Ψ the angle between $\bar{\mu}_1$ and μ_1, then,

$$\cos\Psi = e_1\left\{\bar{e}_1\left[e_1 + \frac{e_p a^2}{\varrho_p^2} + \frac{e_{p+2}a^2}{\varrho_{p+1}^2}\right]\right\}^{-1/2}. \tag{3.6}$$

From this equation it follows that: *The angle between the tangents to C and $\bar{C}$ is constant if and only if,*

$$\frac{e_p}{\varrho_p^2} + \frac{e_{p+2}}{\varrho_{p+1}^2} = \text{constant}. \tag{3.7}$$

Taking the scalar product of each side of (3.4) by μ_{p+2} and calling Θ the angle between $\bar{\mu}_1$ and μ_{p+2} we have,

$$\cos\Theta = \frac{e_{p+2}a}{\varrho_{p+1}}\left\{\bar{e}_1\left[e_1 + \frac{e_p a^2}{\varrho_p^2} + \frac{e_{p+2}a^2}{\varrho_{p+1}^2}\right]\right\}^{-1/2}. \tag{3.8}$$

Since it is assumed that a is different from zero it follows that $1/\varrho_{p+1}$ must equal zero if this expression is to vanish, hence: *A necessary and sufficient condition that the tangent to $\bar{C}$ shall be orthogonal to the $p+1$st normal of C is that C shall lie in at most an S_{p+1}.*

By the Frenet Formulas,

$$\frac{\bar{e}_2}{\bar{\varrho}_1}\bar{\mu}_2 = \frac{d\bar{\mu}_1}{d\bar{s}}. \tag{3.9}$$

Replacing $\bar{\mu}_1$ by the right hand side of (3.4) and performing the differentiation we obtain,

$$
\begin{aligned}
\frac{\bar{e}_2}{\bar{\varrho}_1}\bar{\mu}_2 = {} & \mu_1 \frac{1}{\Omega^2}\left[\frac{e_p a^2 \varrho_p'}{\varrho_p^3} + \frac{e_{p+2} a^2 \varrho_{p+1}'}{\varrho_{p+1}^3}\right] \\
& + \mu_2 \frac{1}{\Omega}\left[\frac{e_2}{\varrho_1}\right] + \mu_{p-1}\frac{1}{\Omega}\left[\frac{e_{p-1} e_p a}{\varrho_{p-1}\varrho_p}\right] \\
& + \mu_p \frac{1}{\Omega}\left[\frac{1}{\Omega}\left(-\frac{e_p a}{\varrho_p}\right)\left\{\frac{e_p a^2 \varrho_p'}{\varrho_p^3} + \frac{e_{p+2} a^2 \varrho_{p+1}'}{\varrho_{p+1}^3}\right\} + \frac{e_p a \varrho_p'}{\varrho_p^2}\right] \\
& + \mu_{p+1}\frac{1}{\Omega}\left[-\frac{e_p e_{p+1} a}{\varrho_p^2} - \frac{e_{p+1} e_{p+2} a}{\varrho_{p+1}^2}\right] \\
& + \mu_{p+2}\frac{1}{\Omega}\left[\frac{1}{\Omega}\left(\frac{e_{p+2} a}{\varrho_{p+1}}\right)\left\{\frac{e_p a^2 \varrho_p'}{\varrho_p^3} + \frac{e_{p+2} a^2 \varrho_{p+1}'}{\varrho_{p+1}^3}\right\} - \frac{e_{p+2} a \varrho_{p+1}'}{\varrho_{p+1}^2}\right] \\
& + \mu_{p+3}\frac{1}{\Omega}\left[\frac{e_{p+2} e_{p+3} a}{\varrho_{p+1}\varrho_{p+2}}\right].
\end{aligned}
\tag{3.10}
$$

Since $\bar{\mu}_2$ is orthogonal to $\bar{\mu}_{p+1}$ and $\bar{\mu}_{p+1} \equiv \mu_{p+1}$, the coefficient of μ_{p+1} in (3.10) must equal zero. That is

$$
a\left(\frac{e_p}{\varrho_p^2} + \frac{e_{p+2}}{\varrho_{p+1}^2}\right) = 0. \tag{3.11}
$$

Hence: *Having given a curve C, if there exists a curve $\bar{C}$ which has for its pth normal the pth normal of C, the pth and $p+1$st curvatures of C must be equal in absolute value and e_p and e_{p+2} must differ in sign.*

It is seen from equation (3.11) that if the space under consideration is Euclidean, a must equal zero and the curves are coincident.

From equations (3.6) and (3.11) we have: *A necessary and sufficient condition that the angle between the tangents to C and $\bar{C}$ shall be constant is that the pth curvature of C shall be constant.*

If $1/\varrho_p$ is constant, $1/\varrho_{p+1}$ is constant and the coefficients of μ_1, μ_p, μ_{p+1}, and μ_{p+2} in equation (3.10) are all zero. Hence: *If the pth curvature of C is constant, the principal normal of $\bar{C}$ is normal to the S_4 determined by the tangent, $p-1$st, pth, and $p+1$st normals of C.*

Since it was assumed that the pth curvature of C was different from zero, we have from equation (3.11) and the theorem immediately following equation (3.8): *The tangent to $\bar{C}$ cannot be parallel to the $p+1$st normal of C.*

If the space under consideration is Euclidean, we have in equation (3.11),

$$
e_p = e_{p+2} = +1. \tag{3.12}
$$

In this case we have: *If two distinct curves have a common pth normal they lie in at most an S_p.*

Making use of the Frenet Formulas, this reduces to,

$$\bar{\mu}_1 = \left[\mu_1 + a\left(-\frac{e_p}{\varrho_p}\mu_p + \frac{e_{p+2}}{\varrho_{p+1}}\mu_{p+2}\right) + a'\mu_{p+1}\right]\frac{ds}{d\bar{s}}. \tag{3.2}$$

But $\bar{\mu}_1$ is orthogonal to $\bar{\mu}_{p+1}$ and $\bar{\mu}_{p+1} \equiv \mu_{p+1}$, hence $a' = 0$ and we have,

$$a = \text{constant}. \tag{3.3}$$

If two curves C and $\bar{C}$ have a common pth normal, they cut off on this normal a segment of constant length.

By (3.3), equation (3.2) reduces to,

$$\bar{\mu}_1 = \left[\mu_1 - \frac{e_p a}{\varrho_p}\mu_p + \frac{e_{p+2}a}{\varrho_{p+1}}\mu_{p+2}\right]\frac{ds}{d\bar{s}}. \tag{3.4}$$

Taking the scalar product of each side of this equation by itself and solving for $ds/d\bar{s}$ we have,

$$\frac{ds}{d\bar{s}} = \left\{\bar{e}_1\left[e_1 + \frac{e_p a^2}{\varrho_p^2} + \frac{e_{p+2}a^2}{\varrho_{p+1}^2}\right]\right\}^{-1/2} \equiv \Omega^{-1/2} > 0. \tag{3.5}$$

If we take the scalar product of each side of (3.4) by μ_1 and call ψ the angle between $\bar{\mu}_1$ and μ_1, then,

$$\cos\psi = e_1\left\{\bar{e}_1\left[e_1 + \frac{e_p a^2}{\varrho_p^2} + \frac{e_{p+2}a^2}{\varrho_{p+1}^2}\right]\right\}^{-1/2}. \tag{3.6}$$

From this equation it follows that: *The angle between the tangents to C and $\bar{C}$ is constant if and only if,*

$$\frac{e_p}{\varrho_p^2} + \frac{e_{p+2}}{\varrho_{p+1}^2} = \text{constant}. \tag{3.7}$$

Taking the scalar product of each side of (3.4) by μ_{p+2} and calling Θ the angle between $\bar{\mu}_1$ and μ_{p+2} we have,

$$\cos\Theta = \frac{e_{p+2}a}{\varrho_{p+1}}\left\{\bar{e}_1\left[e_1 + \frac{e_p a^2}{\varrho_p^2} + \frac{e_{p+2}a^2}{\varrho_{p+1}^2}\right]\right\}^{-1/2}. \tag{3.8}$$

Since it is assumed that a is different from zero it follows that $1/\varrho_{p+1}$ must equal zero if this expression is to vanish, hence: *A necessary and sufficient condition that the tangent to $\bar{C}$ shall be orthogonal to the $p+1$st normal of C is that C shall lie in at most an S_{p+1}.*

By the Frenet Formulas,

$$\frac{\bar{e}_2}{\bar{\varrho}_1}\bar{\mu}_2 = \frac{d\bar{\mu}_1}{d\bar{s}}. \tag{3.9}$$

Replacing $\bar{\mu}_1$ by the right hand side of (3.4) and performing the differentiation we obtain,

$$
\begin{aligned}
\frac{\bar{e}_2}{\bar{\varrho}_1}\bar{\mu}_2 = {} & \mu_1 \frac{1}{\Omega^2}\left[\frac{e_p a^2 \varrho'_p}{\varrho_p^3} + \frac{e_{p+2} a^2 \varrho'_{p+1}}{\varrho_{p+1}^3}\right] \\
& + \mu_2 \frac{1}{\Omega}\left[\frac{e_2}{\varrho_1}\right] + \mu_{p-1}\frac{1}{\Omega}\left[\frac{e_{p-1} e_p a}{\varrho_{p-1}\varrho_p}\right] \\
& + \mu_p \frac{1}{\Omega}\left[\frac{1}{\Omega}\left(-\frac{e_p a}{\varrho_p}\right)\left\{\frac{e_p a^2 \varrho'_p}{\varrho_p^3} + \frac{e_{p+2} a^2 \varrho'_{p+1}}{\varrho_{p+1}^3}\right\} + \frac{e_p a \varrho'_p}{\varrho_p^2}\right] \\
& + \mu_{p+1}\frac{1}{\Omega}\left[-\frac{e_p e_{p+1} a}{\varrho_p^2} - \frac{e_{p+1} e_{p+2} a}{\varrho_{p+1}^2}\right] \\
& + \mu_{p+2}\frac{1}{\Omega}\left[\frac{1}{\Omega}\left(\frac{e_{p+2} a}{\varrho_{p+1}}\right)\left\{\frac{e_p a^2 \varrho'_p}{\varrho_p^3} + \frac{e_{p+2} a^2 \varrho'_{p+1}}{\varrho_{p+1}^3}\right\} - \frac{e_{p+2} a \varrho'_{p+1}}{\varrho_{p+1}^2}\right] \\
& + \mu_{p+3}\frac{1}{\Omega}\left[\frac{e_{p+2} e_{p+3} a}{\varrho_{p+1}\varrho_{p+2}}\right].
\end{aligned}
\tag{3.10}
$$

Since $\bar{\mu}_2$ is orthogonal to $\bar{\mu}_{p+1}$ and $\bar{\mu}_{p+1} \equiv \mu_{p+1}$, the coefficient of μ_{p+1} in (3.10) must equal zero. That is

$$
a\left(\frac{e_p}{\varrho_p^2} + \frac{e_{p+2}}{\varrho_{p+1}^2}\right) = 0. \tag{3.11}
$$

Hence: *Having given a curve C, if there exists a curve $\bar{C}$ which has for its pth normal the pth normal of C, the pth and $p+1$st curvatures of C must be equal in absolute value and e_p and e_{p+2} must differ in sign.*

It is seen from equation (3.11) that if the space under consideration is Euclidean, a must equal zero and the curves are coincident.

From equations (3.6) and (3.11) we have: *A necessary and sufficient condition that the angle between the tangents to C and $\bar{C}$ shall be constant is that the pth curvature of C shall be constant.*

If $1/\varrho_p$ is constant, $1/\varrho_{p+1}$ is constant and the coefficients of μ_1, μ_p, μ_{p+1}, and μ_{p+2} in equation (3.10) are all zero. Hence: *If the pth curvature of C is constant, the principal normal of $\bar{C}$ is normal to the S_4 determined by the tangent, $p-1$st, pth, and $p+1$st normals of C.*

Since it was assumed that the pth curvature of C was different from zero, we have from equation (3.11) and the theorem immediately following equation (3.8): *The tangent to $\bar{C}$ cannot be parallel to the $p+1$st normal of C.*

If the space under consideration is Euclidean, we have in equation (3.11),

$$
e_p = e_{p+2} = +1. \tag{3.12}
$$

In this case we have: *If two distinct curves have a common pth normal they lie in at most an S_p.*

4. Pairs of curves for which the principal normal of one lies in the osculating plane of the other. If the principal normal of $\bar{C}$ lies in the osculating plane of C, the equations of $\bar{C}$ are,

$$\bar{x}(\bar{s}) = x(s) + b\mu_1 + a\mu_2. \tag{4.1}$$

It is assumed that s and $\bar{s}$ are the arc lengths of C and $\bar{C}$ respectively, that a and b are scalar functions of s and are not both zero and that $ds/d\bar{s} > 0$. It is further assumed that,

$$\bar{\mu}_2 = \alpha\mu_1 + \beta\mu_2, \tag{4.2}$$

where,

$$e_1\alpha^2 + e_2\beta^2 = \bar{e}_2. \tag{4.3}$$

Differentiating (4.1) with respect to $\bar{s}$ and simplifying by the Frenet Formulas,

$$\bar{\mu}_1 = \left[\left(1 + b' - \frac{e_1 a}{\varrho_1}\right)\mu_1 + \left(\frac{b e_2}{\varrho_1} + a'\right)\mu_2 + \left(\frac{e_3 a}{\varrho_2}\right)\mu_3\right]\frac{ds}{d\bar{s}}. \tag{4.4}$$

Taking the scalar product of each side of (4.4) by itself and solving for $ds/d\bar{s}$,

$$\frac{ds}{d\bar{s}} = \left\{\bar{e}_1\left[\left(1 + b' - \frac{e_1 a}{\varrho_1}\right)^2 e_1 + \left(\frac{e_2 b}{\varrho_1} + a'\right)^2 e_2 + \frac{e_3 a^2}{\varrho_2^2}\right]\right\}^{-1/2} \equiv \Gamma^{-1/2} > 0. \tag{4.5}$$

If we differentiate (4.4) with respect to $\bar{s}$ and simplify, we have

$$\begin{aligned}
\frac{\bar{e}_2}{\varrho_2}\bar{\mu}_2 = {} & \mu_1\left\{\frac{1}{\Gamma}\left[b'' - \frac{e_1\varrho_1 a'}{\varrho_1^2} + \frac{e_1 a\varrho_1'}{\varrho_1^2} - \frac{e_1 e_2 b}{\varrho_1^2} - \frac{e_1 a'}{\varrho_1}\right] + \frac{\Gamma'}{2\Gamma^2}\left[1 + b' - \frac{e_1 a}{\varrho_1}\right]\right\} \\
& + \mu_2\left\{\frac{1}{\Gamma}\left[\frac{e_2}{\varrho_1}\left(1 + b' - \frac{e_1 a}{\varrho_1}\right) + \left(\frac{e_2\varrho_1 b'}{\varrho_1^2} - \frac{e_2 b\varrho_1'}{\varrho_1^2} + a''\right) - \left(\frac{e_2 e_3 a}{\varrho_2^2}\right)\right] - \frac{\Gamma'}{2\Gamma^2}\left[\frac{e_2 b}{\varrho_1} + a'\right]\right\} \\
& + \mu_3\left\{\frac{1}{\Gamma}\left[\frac{e_3}{\varrho_2}\left(\frac{e_2 b}{\varrho_1} + a'\right) + \frac{e_3\varrho_2 a'}{\varrho_2^2} - \frac{e_3 a\varrho_2'}{\varrho_2^2}\right] - \frac{\Gamma'}{2\Gamma^2}\left[\frac{e_3 a}{\varrho_2}\right]\right\} \\
& + \mu_4\left\{\frac{1}{\Gamma}\left[\frac{e_3 e_4 a}{\varrho_2\varrho_3}\right]\right\}.
\end{aligned} \tag{4.6}$$

In this equation the coefficients of μ_3 and μ_4 must each equal zero. Hence,

$$\begin{cases} \text{(A)} \quad \dfrac{e_2 b}{\varrho_1} + 2a' - \dfrac{a\varrho_2'}{\varrho_2} - \dfrac{\Gamma'}{2\Gamma}a = 0. \\ \text{(B)} \quad \dfrac{1}{\varrho_3} = 0. \end{cases} \tag{4.7}$$

Since $1/\varrho_3 = 0$ we have: *If there exists a curve $\bar{C}$ whose principal normal lies in the osculating plane of a curve C, the curve C lies in at most an S_3.*

A necessary and sufficient condition that C and $\bar{C}$ shall have a common osculating plane is that $\bar{\mu}_1$ and $\bar{\mu}_2$ shall each be expressible as linear combinations of μ_1 and μ_2. From equation (4.4) we see that for this to be the case it is necessary that $1/\varrho_2 = 0$ or that $a = 0$. But if $1/\varrho_2 = 0$, we see that the coefficients of μ_3 and μ_4 in (4.6) are zero and the condition is sufficient. From this it follows that: *Given a curve C, a necessary and sufficient condition that a curve $\bar{C}$ exists whose osculating plane is the osculating plane of C, is that C shall be a plane curve.*

If we take the scalar product of each side of (4.4) by μ_1 and call Δ the angle between the tangents to C and $\bar{C}$,

$$\cos \Delta = \left(1 + b' - \frac{e_1 a}{\varrho_1}\right) e_1. \tag{4.8}$$

$\cos \Delta$ is zero, if and only if,

$$1 + b' - \frac{e_1 a}{\varrho_1} = 0. \tag{4.9}$$

That is, if,

$$b = \int \left(\frac{e_1 a}{\varrho_1} - 1\right) ds + \text{constant}. \tag{4.10}$$

Hence: *A necessary and sufficient condition that the tangents to C and $\bar{C}$ shall be orthogonal is that b shall be expressible in terms of a and ϱ_1 by an integral of the form,*

$$b = \int \left(\frac{e_1 a}{\varrho_1} - 1\right) ds + \text{constant}.$$

From this theorem and from equation (4.4) we have as a corollary: *If the tangent to $\bar{C}$ is orthogonal to the tangent to C at corresponding points, the tangent to $\bar{C}$ is parallel to the plane of the principal normal and binormal of C.*

From equation (4.4) we have still further: *The tangents to C and $\bar{C}$ are parallel if and only if C is a plane curve whose curvature $1/\varrho_1 = -a'/e_2 b$.*

5. **Pairs of curves for which the principal normal of one lies in the plane of the principal normal and binormal of the other.** If the principal normal of $\bar{C}$ lies in the plane of the principal normal and binormal of C, the equations of $\bar{C}$ are,

$$\bar{x}(\bar{s}) = x(s) + a\,\mu_2 + c\,\mu_3, \tag{5.1}$$

where s and $\bar{s}$ are again the arc lengths of C and $\bar{C}$ and where a and c are scalar functions of s, are not both zero and are positive in the directions of the centers of first and second curvatures of C respectively.

Differentiating (5.1) with respect to $\bar{s}$ we obtain,

$$\bar{\mu}_1 = \left[\mu_1 + a\left(-\frac{e_1}{\varrho_1}\mu_1 + \frac{e_3}{\varrho_2}\mu_3\right) + a'\mu_2 + c\left(-\frac{e_2}{\varrho_2}\mu_2 + \frac{e_4}{\varrho_3}\mu_4\right) + c'\mu_3\right]\frac{ds}{d\bar{s}}. \tag{5.2}$$

This reduces to,

$$\bar{\mu}_1 = \left[\left(1 - \frac{a\,e_1}{\varrho_1}\right)\mu_1 + \left(a' - \frac{e_2\,c}{\varrho_2}\right)\mu_2 + \left(\frac{a\,e_3}{\varrho_2} + c'\right)\mu_3 + \left(\frac{e_4\,c}{\varrho_3}\right)\mu_4\right]\frac{ds}{d\bar{s}}. \tag{5.3}$$

If we take the scalar product of each side of (5.3) by itself and solve for $ds/d\bar{s}$ we have,

$$\frac{ds}{d\bar{s}} = \left\{\bar{e}_1\left[\left(1 - \frac{a\,e_1}{\varrho_1}\right)^2 e_1 + \left(a' - \frac{e_2\,c}{\varrho_2}\right)^2 e_2 + \left(\frac{a\,e_3}{\varrho_2} + c'\right)^2 e_3 + \frac{e_4\,c^2}{\varrho_3^2}\right]\right\}^{-1/2} \equiv \Pi^{-1/2}. \tag{5.4}$$

It is assumed that $ds/d\bar{s} > 0$.

The condition for the orthogonality of the tangents to C and $\bar{C}$ at corresponding points is that the coefficient of μ_1 in (5.3) shall be zero. That is, that $e_1/\varrho_1 = 1/a$. But ϱ_1, is the radius of first curvature of C and a is the projection upon μ_2 of the distance between corresponding points on C and $\bar{C}$. Hence we have: *A necessary and sufficient condition that the tangents to C and $\bar{C}$ at a pair of corresponding points shall be orthogonal is that the projection of the distance between the curves at the points, upon the principal normal of C at the point, shall equal the radius of first curvature of C at that point.*

Differentiating both sides of (5.3) with respect to $\bar{s}$ and simplifying by means of the Frenet Formulas,

$$\begin{aligned}\frac{\bar{e}_2}{\bar{\varrho}_1}\bar{\mu}_2 = {} & \mu_1\left\{\frac{1}{\Pi}\left[\frac{e_1\,a\,\varrho_1' - e_1\,\varrho_1\,a'}{\varrho_1^2} + \frac{e_1\,e_2\,c}{\varrho_1\,\varrho_2} - \frac{e_1\,a'}{\varrho_1}\right] - \frac{\Pi'}{2\Pi^2}\left[1 - \frac{a\,e_1}{\varrho_1}\right]\right\} \\ & + \mu_2\left\{\frac{1}{\Pi}\left[\left(1 - \frac{a\,e_1}{\varrho_1}\right)\frac{e_2}{\varrho_1} + a'' - \frac{e_2\,c'\,\varrho_2 - e_2\,c\,\varrho_2'}{\varrho_2^2} - \frac{e_2\,e_3\,a}{\varrho_2^2} - \frac{e_2\,c'}{\varrho_2}\right] - \frac{\Pi'}{2\Pi^2}\left[a' - \frac{e_2\,c}{\varrho_2}\right]\right\} \\ & + \mu_3\left\{\frac{1}{\Pi}\left[\frac{e_3\,a'}{\varrho_2} - \frac{e_2\,e_3\,c}{\varrho_2^2} + \frac{e_3\,a'\,\varrho_2 - e_3\,a\,\varrho_2'}{\varrho_2^2} + c'' - \frac{e_3\,e_4\,c}{\varrho_3^2}\right] - \frac{\Pi'}{2\Pi^2}\left[\frac{a\,e_3}{\varrho_2} + c'\right]\right\}\end{aligned} \tag{5.5}$$

$$+\mu_4\left\{\frac{1}{\Pi}\left[\frac{e_3e_4a}{\varrho_2\varrho_3}+\frac{e_4c'}{\varrho_3}+\frac{e_4c'\varrho_3-e_4c\varrho_3'}{\varrho_3^2}\right]-\frac{\Pi'}{2\Pi^2}\left[\frac{e_4c}{\varrho_3}\right]\right\}$$
$$+\mu_5\left\{\frac{1}{\Pi}\left[\frac{e_4e_5c}{\varrho_3\varrho_4}\right]\right\}.$$

Since μ_2 lies in the plane of the principal normal and binormal of C, the coefficients of μ_1, μ_4, and μ_5 in (5.5) must be zero. Since $1/\Pi \neq 0$ we have the three equations:

$$\text{(A)}\quad \frac{e_1a\varrho_1'}{\varrho_1^2}-\frac{e_1a'}{\varrho_1}+\frac{e_1e_2c}{\varrho_1\varrho_2}-\frac{e_1a'}{\varrho_1}-\frac{\Pi'}{2\Pi}\left(1-\frac{ae_1}{\varrho_1}\right)=0,$$

$$\text{(5.6) (B)}\quad \frac{1}{\varrho_3}\left[\frac{e_3e_4a}{\varrho_2}+e_4c'+\frac{e_4c'}{\varrho_3}-\frac{e_4c\varrho_3'}{\varrho_3^2}-\frac{\Pi'e_4c}{2\Pi}\right]=0,$$

$$\text{(C)}\quad \frac{c}{\varrho_3\varrho_4}=0.$$

From equation (5.6) (C) we see that either $1/\varrho_4 = 0$ or $c = 0$. Let it be assumed that $1/\varrho_4 \neq 0$ and $c = 0$. In this case $1/\varrho_3 \neq 0$. Applying these conditions to (5.6) (B) we find that $a = 0$. But a and c cannot both be zero as this is contrary to our hypotheses. Therefore $1/\varrho_4$ must be zero. Hence: *If a curve $\bar{C}$ exists whose principal normal lies in the plane of the principal normal and binormal of a curve C, C lies in at most an S_4.*

UNIVERSITY OF WISCONSIN, EXTENSION DIVISION, MILWAUKEE.

THE AREA AND BOUNDARY OF MINIMAL SURFACES.[1]

By E. F. Beckenbach.[2]

Let the real analytic functions

$$x_j = x_j(u, v), \qquad j = 1, 2, 3, \tag{1}$$

map the interior of the circle Q with radius R and center (u_0, v_0) upon a surface D. In order that the map be conformal, it is necessary and sufficient that the parameters be isothermic; that is, that

$$E = G, \qquad F = 0, \tag{2}$$

where

$$E = \sum_{j=1}^{3} \left(\frac{\partial x_j}{\partial u}\right)^2, \qquad F = \sum_{j=1}^{3} \left(\frac{\partial x_j}{\partial u}\right)\left(\frac{\partial x_j}{\partial v}\right), \qquad G = \sum_{j=1}^{3} \left(\frac{\partial x_j}{\partial v}\right)^2.$$

Let this choice of parameters be made; then by a well-known theorem of Weierstrass, a necessary and sufficient condition that the surface D be minimal is that the functions (1) be harmonic. This being so, these functions are the real parts of analytic functions,

$$x_j = \Re f_j(z), \qquad z = u + iv.$$

Because of (2), the elements of area and length on the surface reduce respectively to

$$E\,du\,dv \text{ and } E^{1/2}|dz|.$$

We have

$$\begin{aligned} E &= \frac{1}{2}(E+G) \\ &= \frac{1}{2}\sum_{j=1}^{3}\left(\frac{\partial x_j}{\partial u} - i\frac{\partial x_j}{\partial v}\right)\left(\frac{\partial x_j}{\partial u} + i\frac{\partial x_j}{\partial v}\right) \\ &= \frac{1}{2}\sum_{j=1}^{3}|f_j'(z)|^2. \end{aligned}$$

Theorem 1. *If the interior of the circle Q is mapped conformally on a minimal surface of area A, then*

$$A \geqq \pi R^2 E_0, \tag{3}$$

where E_0 is the area magnification ratio of the mapping at (u_0, v_0). In particular, A may be infinite. The equality holds if and only if the map is a circle whose center is the image of (u_0, v_0).

[1] Received January 22, 1932.

[2] National Research Fellow.

This is a generalization of a theorem on complex variables;[3] the present proof follows the same method.

Let the map be given by the isothermic harmonic functions (1). We can suppose, without restricting the generality, that

$$(u_0, v_0) = (0, 0).$$

Let Q' be the circle

$$|z| \leqq R' < R.$$

Q' is mapped on a surface of finite area A', given by

$$A' = \frac{1}{2} \int\int_{Q'} \sum_{j=1}^{3} |f_j'|^2 \, du \, dv.$$

We have

$$f_j(z) = \sum_{t=0}^{\infty} a_{j,t} z^t,$$

the series converging when $|z| < R$. Now

$$|f_j'|^2 = \left(\sum_{t=1}^{\infty} t \, a_{j,t} z^{t-1} \right) \left(\sum_{t=1}^{\infty} t \, \overline{a}_{j,t} \, \overline{z}^{t-1} \right).$$

Putting the double integral into polar coördinates, we have

$$A' = \frac{1}{2} \int_0^{R'} r \, dr \int_0^{2\pi} \sum_{j=1}^{3} \left(\sum_{t=1}^{\infty} t \, a_{j,t} r^{t-1} e^{(t-1)i\theta} \right) \left(\sum_{t=1}^{\infty} t \, \overline{a}_{j,t} r^{t-1} e^{-(t-1)i\theta} \right) d\theta. \tag{4}$$

We now multiply the two series and integrate term by term. When the integer t is different from zero, we have

$$\int_0^{2\pi} e^{ti\theta} \, d\theta = 0,$$

whence

$$A' = \pi \int_0^{R'} \sum_{j=1}^{3} \sum_{t=1}^{\infty} t^2 |a_{j,t}|^2 r^{2t-1} \, dr = \frac{\pi}{2} \sum_{j=1}^{3} \sum_{t=1}^{\infty} t |a_{j,t}|^2 R'^{2t}.$$

When R' approaches R, the sum of this series of positive terms either becomes infinite or approaches the sum

$$\frac{\pi}{2} \sum_{j=1}^{3} \sum_{t=1}^{\infty} t |a_{j,t}|^2 R^{2t},$$

[3] See L. Bieberbach, *Palermo Rend.*, vol. 38, pp. 98–112 (1914); L. R. Ford, *Automorphic Functions*, p. 167 (1929).

provided the latter series converges. In either case,

$$A \geqq \frac{\pi}{2} \sum_{j=1}^{3} |a_{j,1}|^2 R^2 = \pi R^2 E_0.$$

The equality holds if and only if

(5) $$a_{j,t} = 0, \qquad j = 1, 2, 3;\ t > 1,$$

in which case

(6) $$f_j = a_{j,0} + a_{j,1} z.$$

That this map is a circle whose center is the image of (u_0, v_0) becomes apparent when we make the x_1- and x_2-axes tangent, with preservation of positive directions, to the parametric curves, on the surface, given by $v = v_0$ and $u = u_0$ respectively. We can make such a choice without changing the form of (6) because a transformation of the rectangular coördinates does not alter the conformal character of the map. Setting

$$a_{j,1} = b_j - i c_j,$$

we have

$$\frac{\partial x_1}{\partial u} = b_1 = a_{1,1}, \qquad \frac{\partial x_1}{\partial v} = c_1 = 0,$$

$$\frac{\partial x_2}{\partial u} = b_2 = 0, \qquad \frac{\partial x_2}{\partial v} = c_2 = i a_{2,1},$$

$$\frac{\partial x_3}{\partial u} = b_3 = 0, \qquad \frac{\partial x_3}{\partial v} = c_3 = 0.$$

Since the parameters are isothermic, and since b_1 and c_2 are positive,

$$b_1 = c_2 = E^{1/2} = E_0^{1/2},$$

so that

$$x_1 = E_0^{1/2} u, \qquad x_2 = E_0^{1/2} v, \qquad x_3 = 0.$$

Consequently, the circle

$$u^2 + v^2 \leqq R^2$$

is mapped on the circle

$$x_1^2 + x_2^2 \leqq E_0 R^2,$$

the center of the first being mapped on the center of the second.

We note the following four corollaries, and remark once for all that the equalities in all but the second hold if and only if the map is a circle whose center is the image of (u_0, v_0).

1) *If the area at the center (u_0, v_0) of Q is unaltered in magnitude in the above mapping, so that $E_0 = 1$, then the area of the map equals or exceeds that of Q,*

$$A \geqq \pi R^2.$$

2) *If* $R \geqq 1$, *and if there exists a* t *such that*

$$\sum_{j=1}^{3} |a_{j,t}|^2 \geqq \frac{2}{t},$$

then the area of the map equals or exceeds that of Q.

For we have proved incidentally that

$$A \geqq \frac{\pi}{2} \sum_{j=1}^{3} t |a_{j,t}|^2 R^{2t}.$$

3)[4] *In the above mapping, the area* I *of the map of a ring, concentric with the given circle, of radii* R_1 *and* R_2, *where* $0 \leqq R_1 < R_2 \leqq R$, *satisfies the inequality*

$$I \geqq \pi (R_2^2 - R_1^2) E_0.$$

The proof of this corollary consists of changing the limits of integration with respect to r in (4) from $(0, R')$ to (R_1, R_2), or, if $R_2 = R$, to (R_1, R').

As in the first corollary, if $E_0 = 1$, then the area of the map of the ring equals or exceeds that of the ring.

4)[5] *If we denote by* C_1 *and* C_2 *the areas of circles concentric with the given circle, of radii* R_1 *and* R_2 *respectively, where* $0 < R_1 < R_2 \leqq R$, *and denote by* A_1 *and* A_2 *respectively the areas of the maps of these circles, then*

$$\frac{A_2}{A_1} \geqq \frac{C_2}{C_1}.$$

Since the functions are analytic in the larger circle, A_1 is finite. If A_2 is infinite, the inequality holds at once. Otherwise,

$$A_k = \frac{\pi}{2} \sum_{j=1}^{3} \sum_{t=1}^{\infty} t |a_{j,t}|^2 R_k^{2t},$$

$$C_k = \pi R_k^2, \qquad k = 1, 2.$$

Then

$$\frac{A_2}{A_1} - \frac{C_2}{C_1} = \frac{R_2^2 \sum_{j=1}^{3} \sum_{t=2}^{\infty} t |a_{j,t}|^2 (R_2^{2t-2} - R_1^{2t-2})}{\sum_{j=1}^{3} \sum_{t=1}^{\infty} t |a_{j,t}|^2 R_1^{2t}} \geqq 0.$$

The equality holds if and only if equations (5) hold.

THEOREM 2. *If a minimal surface* S *of area* A *is bounded by a curve* B *of length* L, *then*

$$A \leqq \frac{1}{4\pi} L^2. \tag{7}$$

If L *is finite, the equality holds if and only if* B *is a circle.*

[4] This corollary was suggested by Professor L. R. Ford.

[5] Cf. G. Pólya und G. Szegö, *Aufgaben und Lehrsätze*, vol. II, p. 14, § 81 (1925).

If B is a plane curve, the minimal surface lies on the plane and (7) expresses the known isoperimetric property of the circle.

This theorem has been proved by Carleman[6], who showed first that for any function $f(z)$ analytic in Q,

$$\frac{1}{4\pi}\left[\int_0^{2\pi} |f'(R'e^{i\theta})|\, R'\, d\theta\right]^2 \geqq \int_0^{R'} r\, dr \int_0^{2\pi} |f'(z)|^2\, d\theta, \tag{8}$$

the equality holding if and only if $f(z)$ is a linear or linear fractional function whose pole lies outside Q. By means of this inequality, which proves the theorem if B is a plane curve, Carleman then by devious means established the theorem in 3-space.

The present proof establishes the theorem as an immediate consequence of (8) by means of Minkowski's inequality, and has the additional advantage of yielding an obvious generalization to the case of a minimal surface in euclidean n-space.

Let the isothermic harmonic functions (1) map Q on S. Then Q' is mapped on a subregion S' of S, where the area A' and boundary L' of S' are finite. We have

$$\begin{aligned} A' &= \frac{1}{2}\sum_{j=1}^{3} \int_0^{R'} r\, dr \int_0^{2\pi} |f_j'(z)|^2\, d\theta, \\ L'^2 &= \left[\int_0^{2\pi} \left(\frac{1}{2}\sum_{j=1}^{3} |f_j'(R'e^{i\theta})|^2\right)^{1/2} R'\, d\theta\right]^2. \end{aligned} \tag{9}$$

Applying Minkowski's inequality[7] to (9), we have

$$L'^2 \geqq \frac{1}{2}\sum_{j=1}^{3}\left[\int_0^{2\pi} |f_j'(R'e^{i\theta})|\, R'\, d\theta\right]^2 \tag{10}$$

the equality holding if and only if

$$|f_j'(R'e^{i\theta})| = \lambda_j |\varphi'(R'e^{i\theta})|$$

where the λ_j are constants.

From (8) and (10) we get

$$A' \leqq \frac{1}{4\pi} L'^2. \tag{11}$$

Radó has shown[8] that L' is a monotonically increasing function of R'. Consequently, as R' approaches R, L' approaches L if L is finite, or diverges to infinity if L is infinite. If L is finite, A is finite and A'

[6] T. Carleman, *Mathematische Zeitschrift*, vol. 9, pp. 154–160 (1921).

[7] See, for instance, Pólya und Szegö, *loc. cit.*, p. 56, § 91.

[8] T. Radó, *Annals of Mathematics*, vol. 31, p. 461 (1930).

approaches A. If L is infinite, A' either approaches A, if A is finite, or diverges to infinity if A is infinite. In any case, unless both L and A are infinite so that (7) becomes meaningless, (7) results as the limiting case of (11).

If the equality in (7) holds, then the equalities in (8) and (10) must hold. (We do not consider the case where L is infinite.) If the equality in (8) holds for all j,

$$f_j(z) = \frac{a_j z + b_j}{c_j z + d_j},$$

so that

$$\frac{f_j'(z)}{f_k'(z)} = \frac{(c_k z + d_k)^2}{(c_j z + d_j)^2} \cdot \frac{(a_j d_j - b_j c_j)}{(a_k d_k - b_k c_k)}.$$

For the equality in (10) to hold, we must have

$$\left| \frac{f_j'(R' e^{i\theta})}{f_k'(R' e^{i\theta})} \right| = \text{const.}$$

If both equalities hold,

$$\left| \frac{c_k z + d_k}{c_j z + d_j} \right| = \text{const.},$$

an equality which necessitates that f_j is a linear function of f_k. That is,

$$f_j(z) = w_j \varphi(z) + t_j,$$

where w_j and t_j are complex constants and $\varphi(z)$ is a linear or linear fractional function whose pole lies outside Q. Let

$$\varphi(z) = A(u, v) + iB(u, v), \qquad w_j = l_j - i m_j, \qquad t_j = p_j + i q_j;$$

then

$$x_j = l_j A(u, v) + m_j B(u, v) + p_j.$$

Let the x_1- and x_2-axes be tangent, with preservation of positive directions, to the parametric curves, on the surface, given by $v = v_0$ and $u = u_0$ respectively. Then

$$\left(\frac{\partial x_1}{\partial u}\right)_0 = l_1\left(\frac{\partial A}{\partial u}\right)_0 + m_1\left(\frac{\partial B}{\partial u}\right)_0, \qquad \left(\frac{\partial x_1}{\partial v}\right)_0 = l_1\left(\frac{\partial A}{\partial v}\right)_0 + m_1\left(\frac{\partial B}{\partial v}\right)_0 = 0,$$

$$\left(\frac{\partial x_2}{\partial u}\right)_0 = l_2\left(\frac{\partial A}{\partial u}\right)_0 + m_2\left(\frac{\partial B}{\partial u}\right)_0 = 0, \qquad \left(\frac{\partial x_2}{\partial v}\right)_0 = l_2\left(\frac{\partial A}{\partial v}\right)_0 + m_2\left(\frac{\partial B}{\partial v}\right)_0,$$

$$\left(\frac{\partial x_3}{\partial u}\right)_0 = l_3\left(\frac{\partial A}{\partial u}\right)_0 + m_3\left(\frac{\partial B}{\partial u}\right)_0 = 0, \qquad \left(\frac{\partial x_3}{\partial v}\right)_0 = l_3\left(\frac{\partial A}{\partial v}\right)_0 + m_3\left(\frac{\partial B}{\partial v}\right)_0 = 0,$$

where $(\ )_0$ means that we take the derivative designated at (u_0, v_0).

The function $\varphi(z)$ is such that the Jacobian $J\left(\frac{A, B}{u, v}\right)$ is finite and not zero at (u_0, v_0). Consequently,

$$l_3 = m_3 = 0,$$

so that the map lies in the (x_1, x_2)-plane. The map is given by

$$x_1 + i x_2 = f_1(z),$$

where $f_1(z)$ is a linear or linear fractional function with pole outside Q. Such a function maps the interior of Q in a directly conformal manner on the interior of another circle and is the most general function giving such a map. Consequently, if L is finite, the equality in (7) holds if and only if B is a circle.

THEOREM 3. *If the interior of the circle Q is mapped conformally on a minimal surface the length of whose boundary is L, then*

$$L \geqq 2\pi R E_0^{1/2}, \tag{12}$$

where $E_0^{1/2}$ is the length magnification ratio of the mapping at (u_0, v_0). In particular, L may be infinite. The equality holds if and only if the map is a circle whose center is the image of (u_0, v_0).

We obtain (12) from (3) and (7):

$$\frac{1}{4\pi} L^2 \geqq A \geqq \pi R^2 E_0.$$

As a special case, we note that if lengths are unaltered at (u_0, v_0) in the mapping, so that $E_0^{1/2} = 1$, then the length of the boundary of the map equals or exceeds that of Q.

For analytic functions of a complex variable, we have Theorem 3 in the following form.

The function $f(z)$, analytic within a circle of radius R and center z_0, maps the interior of the circle upon a region the length L of whose boundary satisfies the inequality

$$L \geqq 2\pi R |f'(z_0)|.$$

In particular, L may be infinite. The equality holds if and only if

$$f(z) = a_0 + a_1 z.$$

Finally, we remark that every result given in this paper yields an immediate and trivial generalization to the case of a minimal surface in euclidean n-space.

THE RICE INSTITUTE. PRINCETON UNIVERSITY.

EQUIVALENT CONTINUOUS GROUPS.[1]

By Luther Pfahler Eisenhart.

Two continuous groups of order r in the same number n of variables are said to be *similar*, when there exists a non-singular transformation of the variables by means of which the equations of either group are transformable into the other; we prefer the term *equivalent* and shall use it throughout this paper. Lie[2] has given the conditions under which two groups are equivalent. It is the purpose of this paper to establish these results in another and briefer manner.

1. In the sense of Lie a continuous group of order r, say G_r, expressed in terms of n coördinates $x^1, \cdots, x^n$, is characterized by the symbols

$$X_a f = \xi_a^\alpha(x) \frac{\partial f}{\partial x^\alpha} \qquad \begin{pmatrix} a = 1, \cdots, r \\ \alpha = 1, \cdots, n \end{pmatrix}; \tag{1.1}$$

here, and in what follows, a repeated index in a term (in this case α) indicates that the term is the sum of terms as the index takes on all its values (in this case 1 to r). The fundamental theorem of continuous groups is that a necessary and sufficient condition that $X_a f$ are the symbols of a group is that

$$(X_a, X_b) f \equiv X_a(X_b f) - X_b(X_a f) = c_{ab}^{\ e} X_e f \quad (a, b, e = 1, \cdots, r), \tag{1.2}$$

where the c's are constants satisfying the conditions

$$c_{ab}^{\ e} = -c_{ba}^{\ e}, \tag{1.3}$$

$$c_{ab}^{\ e} c_{ed}^{\ f} + c_{bd}^{\ e} c_{ea}^{\ f} + c_{da}^{\ e} c_{eb}^{\ f} = 0. \tag{1.4}$$

Substituting from (1.1) in (1.2), we obtain

$$\xi_a^\alpha \frac{\partial \xi_b^\beta}{\partial x^\alpha} - \xi_b^\alpha \frac{\partial \xi_a^\beta}{\partial x^\alpha} = c_{ab}^{\ e} \xi_e^\beta. \tag{1.5}$$

We denote by q the generic rank of the matrix

$$M = \| \xi_a^\alpha \|, \tag{1.6}$$

that is, the rank for general values of the x's. Evidently $q \leqq r$. If $q = n$, the group is transitive; if $q < n$, intransitive. If $q < r$, we may without

[1] Received May 25, 1932.

[2] Theorie der Transformationsgruppen, Teubner, Leipzig, 1888, vol. 1, pp. 327–354.

loss of generality assume that the matrix $\|\xi_h^\alpha\|$, for $h = 1, \cdots, q$, is of rank q, and put

$$\xi_p^\alpha = \varphi_p^h \xi_h^\alpha \qquad \begin{pmatrix} h = 1, \cdots, q \\ p = q+1, \cdots, r \end{pmatrix}, \tag{1.7}$$

where the quantities φ_p^h are functions of the x's. Then (1.5) may be written

$$\xi_a^\alpha \frac{\partial \xi_b^\beta}{\partial x^\alpha} - \xi_b^\alpha \frac{\partial \xi_a^\beta}{\partial x^\alpha} = (c_{ab}{}^h + c_{ab}{}^p \varphi_p^h)\, \xi_h^\beta. \tag{1.8}$$

If we take $b = p$ in (1.8) and replace ξ_p^β by its expression (1.7), the result is equivalent, in consequence of equations of the form (1.8) in a and h, to

$$X_a \varphi_p^h = \Phi_{ap}^h, \tag{1.9}$$

where

$$\Phi_{ap}^h = c_{ap}{}^h + c_{ap}{}^s \varphi_s^h - \varphi_p^i (c_{ai}{}^h + c_{ai}{}^s \varphi_s^h) \qquad \begin{pmatrix} a = 1, \cdots, r; \\ h, i = 1, \cdots, q; \\ p, s = q+1, \cdots, r \end{pmatrix}. \tag{1.10}$$

When $q < n$, that is when G_r is intransitive, the equations $X_a f = 0$ form a completely integrable system, and because of (1.7) there are q independent equations in the set, and consequently there are $n-q$ independent solutions $\varphi^\sigma(x)$. Without loss of generality, we may assume that the determinant $\left|\frac{\partial \varphi^\sigma}{\partial x^\tau}\right|$ for $\tau = q+1, \cdots, n$ is not zero. If we effect the transformation

$$x'^\lambda = x^\lambda, \quad x^\sigma = \varphi^\sigma(x) \qquad \begin{pmatrix} \lambda = 1, \cdots, q; \\ \sigma = q+1, \cdots, n \end{pmatrix}, \tag{1.11}$$

and note that the functions ξ_a^α transform as the components of a contravariant vector, that is,

$$\xi_a'^\alpha = \xi_a^\beta \frac{\partial x'^\alpha}{\partial x^\beta}, \tag{1.12}$$

then in the new coördinate system, which we now call x^α (dropping primes), we have

$$\xi_a^\sigma = 0 \qquad \begin{pmatrix} a = 1, \cdots, r; \\ \sigma = q+1, \cdots, n \end{pmatrix}. \tag{1.13}$$

Since in the new coördinate system the matrix $\|\zeta_h^\lambda\|$ for $h = 1, \cdots, q$ is of rank q, a set of functions $\xi_{h\lambda}$ are uniquely defined by

$$\xi_{h\lambda} \xi_h^\mu = \delta_\lambda^\mu, \quad \xi_{h\lambda} \xi_i^\lambda = \delta_{ih} \qquad (\lambda, \mu, i, h = 1, \cdots, q), \tag{1.14}$$

where

$$\delta_\lambda^\mu = 1 \text{ or } 0, \text{ as } \mu = \lambda \text{ or } \mu \neq \lambda,$$
$$\delta_{ih} = 1 \text{ or } 0, \text{ as } i = h \text{ or } i \neq h.$$

In this coördinate system equations (1.8), when a and b take values 1 to q reduce to

$$(1.15)\quad \xi_h^\lambda \frac{\partial \xi_i^\mu}{\partial x^\lambda} - \xi_i^\lambda \frac{\partial \xi_h^\mu}{\partial x^\lambda} = (c_{hi}{}^j + c_{hi}{}^p \varphi_p^j)\, \xi_j^\mu \begin{pmatrix} \lambda, \mu, h, i, j = 1, \cdots, q; \\ p = q+1, \cdots, r \end{pmatrix},$$

and equations (1.9) for $a = 1, \cdots, q$ may be written in the form

$$(1.16)\quad \frac{\partial \varphi_p^h}{\partial x^\lambda} = \Phi_{ip}^h\, \xi_{i\lambda},$$

in consequence of (1.14).

We define functions $L_{\beta\mu}^\alpha$ by

$$(1.17)\quad L_{\beta\mu}^\lambda = \xi_h^\lambda \frac{\partial\, \xi_{h\mu}}{\partial\, x^\beta} = -\xi_{h\mu} \frac{\partial\, \xi_h^\lambda}{\partial\, x^\beta},\ L_{\beta\mu}^\sigma = 0 \begin{pmatrix} \lambda, \mu = 1, \cdots, q; \\ \beta = 1, \cdots, n; \\ \sigma = q+1, \cdots, n \end{pmatrix}.$$

(The reason for this definition will appear in § 2). From (1.17) we have in consequence of (1.14)

$$(1.18)\quad \frac{\partial\, \xi_h^\lambda}{\partial\, x^\beta} + \xi_h^\mu\, L_{\beta\mu}^\lambda = 0,$$

$$(1.19)\quad \frac{\partial\, \xi_{h\lambda}}{\partial\, x^\beta} - \xi_{h\mu}\, L_{\beta\lambda}^\mu = 0.$$

From (1.17) and (1.15) we have

$$(1.20)\quad L_{\mu\nu}^\lambda - L_{\nu\mu}^\lambda = (c_{hi}{}^j + c_{hi}{}^p\, \varphi_p^j)\, \xi_{h\nu}\, \xi_{i\mu}\, \xi_j^\lambda.$$

2. If the coördinates x^i undergo a non-singular transformation, say

$$(2.1)\quad x'^\alpha = \psi^\alpha(x^1, \cdots, x^n),$$

the functions $\xi_a'^i$ are related to the ξ's by (2.1) and

$$(2.2)\quad \xi_a'^\alpha(x') = \xi_a^\beta \frac{\partial\, x'^\beta}{\partial\, x^\alpha}.$$

Consequently a necessary and sufficient condition that two groups G_r and G_r', in the same number of variables x^α and x'^α and with the symbols $\xi_a^\alpha \dfrac{\partial f}{\partial x^\alpha}$ and $\xi_a'^\alpha \dfrac{\partial f}{\partial x'^\alpha}$ be equivalent is that there exist a non-singular transformation such that (2.2) hold. Evidently a necessary condition is that the generic rank of the matrices $\|\xi_a^\alpha\|$ and $\|\xi_a'^\alpha\|$ be equal.

We consider first the case when the generic rank q is equal to r. If $r < n$, the coördinates x^α and x'^α can be chosen so that in accordance with (1.13) we have

$$(2.3)\quad \xi_a^\sigma = \xi_a'^\sigma = 0 \qquad (a = 1, \cdots, r;\ \sigma = q+1, \cdots, n).$$

From (2.2) for $\alpha = r+1, \cdots, n$ we have

$$\frac{\partial x'^{\sigma}}{\partial x^{h}} = 0 \qquad (h = 1, \cdots, q;\ \sigma = q+1, \cdots, n) \tag{2.4}$$

and consequently $x'^{\sigma} = \varphi^{\sigma}(x^{r+1}, \cdots, x^{n})$. Since the φ's are independent, we may take them as new coördinates $x^{r+1}, \cdots, x^{n}$, and not change $x^{1}, \cdots, x^{r}$. In this new coördinate system equations (2.3) hold and in all generality we have

$$x'^{\sigma} = x^{\sigma} \qquad (\sigma = q+1, \cdots, n). \tag{2.5}$$

When α in (2.2) takes the values 1 to $q(=r)$, these equations may be written, in consequence of (1.14), in the form

$$\frac{\partial x'^{\lambda}}{\partial x^{\mu}} = \xi'^{\lambda}_{a}(x')\,\xi_{a\mu}(x) \qquad (\lambda, \mu = 1, \cdots, r). \tag{2.6}$$

If in the functions $\xi'^{\lambda}_{a}(x')$ we replace x'^{σ} by x^{σ} in accordance with (2.5), we have a system of differential equations in the independent variables $x^{1}, \cdots, x^{q}$ and with $x^{q+1}, \cdots, x^{n}$ entering as parameters. By means of (1.18) and (1.19), and similar equations in the ξ''s we obtain from (2.6)

$$\frac{\partial^{2} x'^{\lambda}}{\partial x^{\mu}\,\partial x^{\nu}} = -L'^{\lambda}_{\pi\tau}\,\xi'^{\tau}_{h}\,\xi_{h\mu}\,\xi'^{\pi}_{l}\,\xi_{l\nu} + \xi'^{\lambda}_{h}\,\xi_{h\pi}\,L^{\pi}_{\nu\mu} \quad (\lambda, \mu, \nu, \pi, \tau, h = 1, \cdots, q). \tag{2.7}$$

For the case $q = r$ equations (1.20) are

$$L^{\lambda}_{\mu\nu} - L^{\lambda}_{\nu\mu} = c_{hi}{}^{j}\,\xi_{h\nu}\,\xi_{i\mu}\,\xi^{\lambda}_{j} \qquad (h, i, j = 1, \cdots, q).$$

Because of these relations it follows from (2.7) that the conditions of integrability of equations (2.6) are satisfied identically. Hence a solution is determined by taking for initial values of $x'^{1}, \cdots, x'^{r}$ arbitrary functions of $x^{r+1}, \cdots, x^{n}$. If these are chosen so that the jacobian of the solutions with respect to $x^{1}, \cdots, x^{r}$ is different from zero, these solutions and (2.5) define a non-singular transformation of G_r into G'_r. Hence we have:

Two intransitive groups G_r and G'_r in the same number of variables, with the same constants of structure and whose matrices M are of rank r, are equivalent; the equations of transformation involve r arbitrary functions.

If $r = n$, that is, if the group is simply transitive, we have in place of (2.3), (2.5) and (2.6) only equations (2.6) in which $\lambda, \mu, a = 1, \cdots, n$. As before these equations are completely integrable, and their solution involves n arbitrary constants, the initial values of the x''s. Hence we have:

Two simply transitive groups in the same number of variables and with the same constants of structure are equivalent, and the equations of transformation of one into the other involves n arbitrary constants.

We consider next the case when $q<r$. For the group G_r we choose the ξ's so that $\| \xi_h^\alpha \|$ for $h = 1, \cdots, q$ is of rank q, then the matrix $\| \xi_h'^\alpha \|$ for $h = 1, \cdots, q$ is of rank q, and analogously to (1.7), we put

$$\xi_p'^\alpha = \varphi_p'^h \, \xi_h'^\alpha \qquad (p = q+1, \cdots, r). \tag{2.8}$$

From these equations and (2.2) we have

$$\varphi_p'^h(x') = \varphi_p^h(x) \qquad \begin{pmatrix} h = 1, \cdots, q; \\ p = q+1, \cdots, r \end{pmatrix}. \tag{2.9}$$

There are $q(r-q)$ of these equations. Evidently in order that the two groups be equivalent, it is necessary that these equations be consistent, and that it be impossible to eliminate the x''s and get a relation between the x's, and vice-versa. We shall show that this condition is also sufficient.

If $q<n$, we assume that the coördinates x' and x are chosen so that (2.3) and (2.5) hold. As in the case when $q = r$, we get the system of equations (2.6) in which now $\lambda, \mu, a = 1, \cdots, q$. When in the ξ''s in these equations and in (2.9) we replace x'^σ for $\sigma = q+1, \cdots, n$ by x^σ, we have a mixed system of differential equations in $x'^1, \cdots, x'^q$ in the independent variables $x^1, \cdots, x^q$, which system involves $x^{q+1}, \cdots, x^n$ as parameters. As in the preceding case we have (2.7), and because of (1.20) the conditions of integrability of (2.6) are satisfied because of (2.9).

From (1.16) we have

$$\frac{\partial \varphi_p^h}{\partial x^\lambda} = \Phi_{lp}^h \, \xi_{l\lambda}, \qquad \frac{\partial \varphi_p'^h}{\partial x'^\lambda} = \Phi_{lp}'^h \, \xi_{l\lambda}'. \tag{2.10}$$

If equations (2.9) are differentiated with respect to x^μ, for $\mu = 1, \cdots, q$, the resulting equations are reducible in consequence of (2.5), (2.6) and (2.10) to

$$(\Phi_{lp}'^h - \Phi_{lp}^h) \, \xi_{l\mu} = 0,$$

which are satisfied because of (2.9), as is evident from (1.10).

Thus in applying the general theory of mixed systems[3] to the present case, we see that equations (2.9) are the system F_0, but all the systems $F_1, \cdots$, are consequences of F_0. Hence there exists a solution, if equations (2.9) are consistent, and do not lead to relations between the x's alone or the x''s alone. If they are consistent, s of the functions $\varphi_p'^h$ are independent, say $\varphi'^1, \cdots, \varphi'^s$ and s must be equal to or less than q, because (2.5) also must hold. Consequently the jacobian of these φ''s and x'^σ must be of

[3] Cf. Eisenhart, Non-Riemannian Geometry, American Mathematical Society Colloquium Publications, vol. 8, 1927, p. 17. The notation $F_0, \cdots,$ above is defined there, and references hereafter to equations in this treatment are of the form E (8.5).

rank $n-q+s$, and therefore the rank of $\left\| \frac{\partial \varphi'^{\alpha}}{\partial x'^{\lambda}} \right\|$, for $\alpha = 1, \cdots, s$ and $\lambda = 1, \cdots, q$, is s. Hence the equations $\varphi'^{\alpha} = \varphi^{\alpha}$ can be solved for s of $x'^1, \cdots, x'^q$ as functions of $x'^{q+1}, \cdots, x'^n, x^1, \cdots, x^n$; there is no loss in generality in assuming that these are $x'^1, \cdots, x'^s$. When in these we replace x'^{σ} for $\sigma = q+1, \cdots, n$ by x^{σ}, we have

$$x'^{\alpha} = \psi^{\alpha}(x'^{s+1}, \cdots, x'^{q}; x^1, \cdots, x^n) \qquad (\alpha = 1, \cdots, s). \tag{2.11}$$

These are the analogue of E (8.5)[4] in which $x^{q+1}, \cdots, x^n$ enter as parameters. Then we have to solve the complete system (analogue of E (8.7))

$$\frac{\partial x'^{\varrho}}{\partial x^{\lambda}} = f^{\varrho}(x'^{s+1}, \cdots, x'^{q}; x^1, \cdots, x^n) \qquad \begin{pmatrix} \varrho = s+1, \cdots, q \\ \lambda = 1, \cdots, q \end{pmatrix} \tag{2.12}$$

obtained from (2.6) on replacing $x'^1, \cdots, x'^s$ by the expressions from (2.11) and x'^{σ} by x^{σ} for $\sigma = q+1, \cdots, n$. The solution of (2.12) involves $q-s$ arbitrary functions of the parameters $x^{q+1}, \cdots, x^n$.

If $q = n$, we do not have (2.3) and (2.5), and hence (2.6) and (2.9) involve all the x's as independent variables, but the above process applies just the same.

Hence we have the following theorem established by Lie in a different manner:

Two r-parameter groups in the same number of variables, with the same constants of structure and for which the generic ranks of the matrices $\|\xi\|$ *and* $\|\xi'\|$ *are less than r are equivalent, when, and only when, these respective ranks are equal, and the corresponding set of equations* (2.9) *are consistent and do not lead to a relation between the variables of either set.*

[4] See previous footnote.

ON THE CELLULAR SUBDIVISION OF n-DIMENSIONAL REGIONS.[1]

By Stewart S. Cairns.

1. Introduction. The region $(R+B)$. The problem of dividing an n-dimensional region into the cells of a complex has been treated in the literature only for the case where the boundary is defined by analytic functions.[2] The treatment referred to is inductive with respect to the dimensionality, and, because of the complicated character of the subdivision, is of value chiefly as a demonstration that the region in question is a topological complex.[3] In the present paper, the problem is treated for a region whose boundary is of class C', by a method which leads to a subdivision into very simple cells. The method is a generalization of one employed by W. F. Osgood[4] in the plane, as a basis for a number of fundamental theorems of analysis. O. D. Kellogg[5] makes use of such a subdivision in 3-space in order to establish rigorously Green's and Stokes' Theorems.

Let $(y_1, \cdots, y_n)$ be rectangular cartesian coördinates in euclidean n-space. A closed $(n-1)$-dimensional point-set will be called a *regular topological $(n-1)$-manifold* if the points thereof in the neighborhood of any given point can be represented by giving one of the y's as a single-valued function of class C' of the remaining y's.

This paper deals with an arbitrary limited region, R, of euclidean n-space, whose boundary, B, consists of a finite number of distinct regular $(n-1)$-spreads.

2. The rectangles of the subdivision. A set (K) of distinct i-rectangles $(i = 0, \cdots, p \leqq n)$[6] will be called *a set of related rectangles* if

[1] Received February 25, 1932.—The subdivision theorem of this paper forms part of the writer's thesis, The cellular subdivision and approximation of regular spreads (Harvard, 1931), written under Professor Marston Morse. The theorem appears also in two abstracts with the same title (Bull. Amer. Math. Soc., vol. 36, 1930, p. 481; Proc. Nat. Acad. Sci., vol. 16, 1930, pp. 488–491).

[2] B. L. van der Waerden, Topologische Begründung des Kalküls der abzählenden Geometrie, Math. Ann. 102 (1929), pp. 360–362. Solomon Lefschetz, Topology (1930), Chapter VIII. Koopman and Brown, On the covering of analytic loci by complexes, Trans. Amer. Math. Soc., 34 (1932), pp. 231–251.

[3] These remarks apply also to the problem of subdividing n-manifolds in $(n+r)$-space. The writer plans to publish later the details of his work on the latter problem. For the results, see the abstracts referred to above.

[4] Lehrbuch der Funktionentheorie I (1928), pp. 190–196.

[5] Foundations of potential theory (1929), Ch. IV, § 11.

[6] A 0-rectangle means a point. A 1-rectangle means a segment.

(I) Every i-rectangle ($i = 0, \cdots, p$) of (K) which contains a point of R contains just an i-cell in R.

(II) Every i-rectangle ($i = 1, \cdots, p$) of (K) which meets B contains just an $(i-1)$-cell on B.

(III) The boundary of every q-rectangle ($1 \leqq q \leqq p$) of the set (K) is the sum of i-rectangles ($i = 0, \cdots, q-1$) of the same set, and every i-rectangle of (K) is on the boundary of some q-rectangle of (K).[7]

The primary object of this paper is the construction of a set of related rectangles containing $(R+B)$.

An i-rectangle K_i ($i = 0, \cdots, n$) will be called *of* 0*th type* if, with its boundary, it fails to meet B. An i-rectangle K_i ($i = 1, \cdots, n$) will be called *of first type* if (1) some pair of opposite faces, K_{i-1} and K'_{i-1}, are both of 0th type, (2) K_i plus its boundary contains just an $(i-1)$-cell, σ_{i-1}, and its boundary on B, (3) σ_{i-1}, with its boundary, projects in one-to-one fashion into K_{i-1} plus boundary. Then σ_{i-1} obviously divides K_i into two i-cells. Now let j have one of the values $(2, \cdots, n)$ and suppose that the general i-rectangle K_i ($i = j-1, \cdots, n$) of $(j-1)$th type has been so defined that it cuts from B just an $(i-1)$-cell and is thereby divided into two i-cells. *Then* K_i ($i = j, \cdots, n$) *will be called of* j*th type if*

(1) *It has at least one pair of opposite faces* (K_{i-1}, K'_{i-1}), *where* K_{i-1} *is of* $(j-1)$*th type and* K'_{i-1} *is of* 0*th type.*

(2) K_i *plus its boundary contains just an* $(i-1)$*-cell,* σ_{i-1}, *and its boundary on* B.

(3) *The projection of* σ_{i-1}, *plus boundary, onto* K_{i-1} *coincides with one of the two* $(i-1)$*-cells, plus boundary, into which* K_{i-1} *is divided by* B, *and the projective correspondence thus established is one-to-one.*

It is easily seen that σ_{i-1} *divides* K_i *into two* i*-cells. This completes a definition of* i*-rectangles of* j*th type* ($j = 0, \cdots, n$; $i = j, \cdots, n$). *We shall call* K_{i-1} *a base of* K_i. *When* K_i *is of first type, either of its two faces of* 0*th type may be taken as base.*

The rectangles we employ will have their edges parallel to coördinate axes. If, in the above definition, the y_λ-axis is parallel to the edges of K_i joining K_{i-1} to K'_{i-1}, then K_i will be said to *extend in the* y_λ*-direction.*

3. The radius s. LEMMA. *A positive radius* s *exists, so small that an* $(n-1)$*-sphere of radius not greater than* s *with its center on* B *contains in its interior just an* $(n-1)$*-cell on* B.

By our definition of regular $(n-1)$-manifold (§ 1), the direction of the outer normal to B varies continuously on B. Therefore a positive radius, s',

[7] In other words, the rectangles (K), regarded as cells, constitute a complex [O. Veblen, Analysis Situs (1931), pp. 76, 77]. This implies that the cells mentioned under (I) and (II) also constitute a complex.

may be chosen so small that any two outer normals to B at points within distance s' of one another form an angle less than 45°. Now let S_{n-1} be any $(n-1)$-sphere of radius s' with center, P, on B, and let E_{n-1} be the set of all points which can be connected with P on B inside S_{n-1}. Then any 2-plane, Π_2, containing the normal to B at P meets E_{n-1} in a set of arcs of curves with continuously turning tangents. One of these arcs, γ, passes through P. By the above restriction on s', the tangent to γ at any point makes an angle less than 45° with the tangent at P.[8] Therefore γ projects in one-to-one fashion onto a segment, γ', of the line in which Π_2 meets the tangent $(n-1)$-plane to B at P. As Π_2 turns through all positions containing the normal at P, γ' sweeps out an $(n-1)$-cell, E'_{n-1}, on the tangent $(n-1)$-plane, while γ has E_{n-1} for its locus. But then E_{n-1} corresponds to E'_{n-1} by a one-to-one projection and is accordingly an $(n-1)$-cell. The distance from P to the point-set $(B-E_{n-1})$ is a positive and continuous function of P on B. Its minimum value, s, satisfies the lemma, for an $(n-1)$-sphere about P of radius not greater than s contains no points on $(B-E_{n-1})$ and cuts from E_{n-1} a single $(n-1)$-cell.[9]

COROLLARY. *Let $(a_1, \cdots, a_n)$ be the numerical values of the direction cosines of the normal to B at any point, and let s be chosen so small that none of the a's varies by $\frac{1}{4\sqrt{n}}$ on the $(n-1)$-cell, E_{n-1}, of the lemma. Then, if $a_j \geqq \frac{1}{2\sqrt{n}}$ at the center of the sphere, E_{n-1} projects in one-to-one fashion onto the $(n-1)$-plane $y_j = 0$. Furthermore, if l is a segment with endpoints on E_{n-1}, and l' is the projection of l onto $y_j = 0$, then $l < 4\sqrt{n}\, l'$.*

In view of the restrictions on s, $a_j > \frac{1}{4\sqrt{n}}$ over all E_{n-1}. To prove that the projection of E_{n-1} on $y_j = 0$ is single-valued, let the normal to B at P be replaced, in the proof of the lemma, by a line through P parallel to the y_j-axis, and let the tangent $(n-1)$-plane at P be replaced by $y_j = 0$. In the modified proof, γ is nowhere parallel to the y_j-axis, since $a_j > 0$ all over E_{n-1}. For the rest of the corollary, we need only consider that if a_j equalled $\frac{1}{4\sqrt{n}}$ all over E_{n-1}, we should have $l \leqq 4\sqrt{n}\, l'$. But, by our restrictions, $a_j > \frac{1}{4\sqrt{n}}$ all over E_{n-1}.

[8] For, the line, N, normal to B at P is also the normal to γ in the plane Π_2. The normal to B at any other point, P', on γ, is in the $(n-1)$-plane normal to the tangent, T, to γ at P'. Hence some line in the latter $(n-1)$-plane makes an angle less than 45° with N. Hence the acute angle between T and N exceeds 45°, and the angle between T and the tangent to γ at P is less than 45°.

[9] For it cuts a single arc from each of the curves γ which constitute E_{n-1}.

4. **Rectangles of norm d.** Any k-rectangle ($k = 0, \cdots, n$) of the form

$$(4.1) \qquad \begin{cases} p_i d < y_i < q_i d & (i = i_1, \cdots, i_k), \\ y_j = p_j d & (j \neq i_1, \cdots, i_k), \end{cases}$$

where the p's and q's are integers and d is an arbitrary positive constant will be called a k-rectangle *of norm* d.

LEMMA. *If* $d \leqq \frac{s}{16n}$, *then any point,* P, *at which* $a_\lambda \geqq \frac{1}{2\sqrt{n}}$ *is an interior or boundary point of some* n*-rectangle*

$$K_n \begin{cases} p_i d < y_i < (p_i+1) d \\ p_\lambda d < y < (p_\lambda + q_\lambda) d \end{cases} \qquad (i = 1, \cdots, \lambda-1, \lambda+1, \cdots, n)$$

of maximum diameter less than $\frac{s}{2}$, *where* K_n *is of norm* d, *of first type* (§ 2) *and extends in the* y_λ*-direction* (§ 2).

By § 3, a sphere of radius $\frac{s}{2}$ about P cuts from B an $(n-1)$-cell, E_{n-1}, whose projection on $y_\lambda = 0$ is an $(n-1)$-cell, E'_{n-1}, containing all points within distance $\frac{s}{8\sqrt{n}}$ of the projection, P', of P. But, since $d \leqq \frac{s}{16n}$, the distance $\frac{s}{8\sqrt{n}}$ is over twice the diameter of an $(n-1)$-cube, K_{n-1},

$$(4.2) \qquad \begin{cases} p_i d < y_i < (p_i+1) d \\ y_\lambda = 0 \end{cases} \qquad (i = 1, \cdots, \lambda-1, \lambda+1, \cdots, n)$$

of norm and edge-length d with P' in its interior or on its boundary. Hence K_{n-1} is the projection of an $(n-1)$-cell, σ_{n-1}, on E_{n-1} of maximum diameter $\frac{s}{4}$ or less. But then the projection of σ_{n-1}, plus boundary, onto the y_j-axis is contained in an interval

$$(4.3) \qquad p_\lambda d < y_\lambda < (p_\lambda + q_\lambda) d$$

where $(q_\lambda d) < \left(\frac{s}{4} + 2d\right)$ and (p_λ, q_λ) are integers. If the equation in (4.2) is replaced by the inequality (4.3), the resulting set of inequalities defines an n-rectangle, K_n, which is easily seen to satisfy the lemma.

The n-rectangle K_n will be called an *admissible* n*-rectangle of first type* if the interval (4.3) is the shortest segment of norm d on the y_λ-axis containing the projection of σ_{n-1} plus boundary. Similarly, in the above proof, consider a bounding i-cube K_i ($i = 0, \cdots, n-2$), of K_{n-1}, and let σ_i denote the i-cell on the boundary of σ_{n-1} which projects into K_i.

Let inequality (4.3) now define an interval on the y_λ-axis containing the projection of o_i plus boundary. If we let inequality (4.3) replace the equation $y_\lambda = 0$ in the defining equations and inequalities of K_i, we obtain the definition of an $(i+1)$-rectangle, K_{i+1}, of first type and norm d on the boundary of K_n. If the interval (4.3) is here chosen as small as possible, K_{i+1} will be called an *admissible* $(i+1)$*-rectangle of first type.*

COROLLARY 1. *Let* $K^\lambda(d)$ *denote the set consisting of*

(1) *all the admissible n-rectangles* (K_n) *satisfying the lemma,*

(2) *the various admissible i-rectangles* $(i = 1, \cdots, n-1)$ *of norm d and first type on the boundaries of the rectangles* (K_n), *and*

(3) *the various i-rectangles* $(i = 0, \cdots, n-1)$ *of 0th type, norm and edge-length d, on the boundaries of the rectangles* (K_n).

Then $K^\lambda(d)$ *is a set of related rectangles containing all points on B where* $a_\lambda \geqq \dfrac{1}{2\sqrt{n}}$.

There is no difficulty in verifying for $K^\lambda(d)$ the various properties required by the definition of sets of related rectangles (§ 2). The remainder of the corollary is obvious.

COROLLARY[10] 2. *Given any positive value* $\delta \leqq \dfrac{s}{n}$, *consider an i-rectangle,* $K_i (i = 1, \cdots, n-1)$, *of some norm, d, and of diameter less than* $\dfrac{\delta}{8n}$. *If* (1) K_i, *or its boundary, meets B, and* (2) *within distance* $\dfrac{s}{2}$ *of* K_i *there is a point at which* $a_\lambda \geqq \dfrac{1}{2\sqrt{n}}$, λ *being a value such that no edge of* K_i *is parallel to the* y_λ*-axis, then there exists an* $(i+1)$*-rectangle,* K_{i+1}, *of norm d and first type which extends in the* y_λ*-direction and has* K_i *for a cross-section. The maximum diameter of* K_{i+1} *can be made less than* δ.

COROLLARY 3. *If, in Corollary 2,* K_i *is an i-rectangle of jth type, then* K_i *divides* K_{i+1} *into two* $(i+1)$*-rectangles, both of* $(j+1)$*th type, both extending in the* y_λ*-direction and both having* K_i *for base (§ 2).*

This result is easy to establish, using the definitions in § 2.

If, in Corollary 2, the edge-length of K_{i+1} parallel to the y_λ-axis is as small a multiple of d as possible, K_{i+1} will be called an *admissible* $(i+1)$*-rectangle of cross-section* K_i and norm d. If K_{i+1} satisfies this same condition in Corollary 3, then the two rectangles into which it is divided by K_i will be called *admissible* $(i+1)$*-rectangles of base* K^i *and norm d.*

The following easily verified statements are collected here for purposes of reference.

[10] This corollary differs but slightly from the lemma. Both are convenient for future reference.

(I) *A rectangle of norm d is also of norm d' provided d/d' is a positive integer.*

(II) *Let K_i be any rectangle of norm d and first type which extends in the y_λ-direction and cuts from B a cell at every point of which $a_\lambda \geqq \frac{1}{2\sqrt{n}}$. Then, if d/d' is a positive integer, K_i is the sum of the rectangles of the set $K^\lambda(d')$ which it contains plus a set of rectangles of 0th type, norm d' and edge-length d'.* [See Corollary 1.]

(III) *Let K_{i+1} be an admissible $(i+1)$-rectangle of any base, K_i, and norm d, where K_{i+1} extends in the y_λ-direction and cuts from B a cell at every point of which $a_\lambda \geqq \frac{1}{2\sqrt{n}}$. Suppose K_i plus its boundary consists of a set (K) of distinct rectangles of norm d and of various types. Then $\overline{K}_{i+1}$[11] may be subdivided into* (1) *the rectangles (K);* (2) *those admissible rectangles of norm d with bases in the set (K) which extend in the y_λ-direction and contain points on $\overline{K}_{i+1}$;* (3) *the rectangles of the set $K^\lambda(d)$ which contain points on $\overline{K}_{i+1}$ but not on the rectangles mentioned under* (1) *and* (2); *and* (4) *the rectangles of 0th type, norm d and edge-length d on $\overline{K}_{i+1}$ but not on the rectangles mentioned under* (1), (2), *and* (3). *If (K) is a set of related rectangles, so is the subdivision of $\overline{K}_{i+1}$ just defined.*

5. The rectangles (R^k). An inductive hypothesis. The following hypothesis underlies the remainder of our work.

HYPOTHESIS. *For some value $k<n$, there exists a set (R^k) of related rectangles (§ 2) of some norm, d_k, containing all points on B where one or more of the inequalities*

$$a_j \geqq \frac{1}{\sqrt{n}} \qquad (j = 1, \cdots, k) \tag{5.1}$$

holds. Furthermore, any rectangle of the set (R^k) is of diameter less than $s/2$ and is of ith type $(i \leqq k)$ [§ 2]. If it meets B, it extends in the y_λ-direction (§ 2) for some $\lambda \leqq k$ and lies within distance $s/2$ of some point at which $a_\lambda \geqq \frac{1}{2\sqrt{n}}$.

By § 4, Corollary 1, this hypothesis is satisfied, for $k = 1$, if we let $d_1 = s/16n$ and let $(R^1) = K^1(d_1)$.

6. Subdivisions of the rectangles (R^k). In the next section we shall construct a set (R^{k+1}) of rectangles to satisfy the hypothesis in § 5, read with $(k+1)$ in place of k. As a preliminary, we establish the following lemma and its corollary.

[11] If K is a rectangle, $\overline{K}$ is used to denote K plus its boundary.

LEMMA. *Let $\bar{K}_p$ denote any rectangle of norm d_k plus its boundary. If, for some $\lambda \leqq k$, $\bar{K}_p$ meets B only in points at which $a_\lambda \geqq \frac{1}{2\sqrt{n}}$; then, for any positive δ, it is possible to subdivide $\bar{K}_p$ into a set (K'_p) of rectangles of some norm, d_{k+1}, in such a way that*

(A) *Each of the rectangles (K'_p) is of diameter less than δ.*

(B) *A subset, (K''_p), of (K'_p) subdivides all rectangles of the set (R^k) on $\bar{K}_p$ in such a way that the Hypothesis in § 5 is satisfied by (R^k) after the subdivision, provided d_{k+1} replaces d_k.*

(C) *The rectangles (K'_p), regarded as cells, constitute a complex* [cf. § 2 (III)].

If $p = 1$, the lemma is obvious. If $p > 1$, let the lemma be assumed for the value $p - 1$. It may then be applied to the two faces of K_p perpendicular to the y_λ-axis and also to every cross-section of K_p by one of the planes $y_\lambda = m d_k$, where m is an integer.

(1) Let subdivisions be accordingly made in each of these cases, and let (K'_{p-1}) be the set of all rectangles of these subdivisions. Let every rectangle of (K'_{p-1}) be of diameter less than both $\frac{d_k}{32n}$ and $\frac{\delta}{8n}$. We shall denote by (K''_{p-1}) the subset of (K'_{p-1}) which subdivides rectangles of the set (R^k) in accordance with Part B of the lemma.

Since the rectangles (R^k) are of norm d_k, (K''_{p-1}) is a subdivision of just those rectangles of the set (R^k) on $\bar{K}_p$ which have no edges parallel to the y_λ-axis. Let K_i be any rectangle of the set (K''_{p-1}) which meets B. If K_i is in either face of $\bar{K}_p$ perpendicular to the y_λ-axis, then it is base of just one admissible $(i+1)$-rectangle, K_{i+1} (§ 4, under Corollary 3), on $\bar{K}_p$, where K_{i+1} extends in the y_λ-direction. In any other case, K_i is the base of two such $(i+1)$-rectangles.

(2) Let every such $(i+1)$-rectangle[12] be constructed.

Again consider K_i. If K_i is on the boundary of any rectangle, K_j, which belongs to the set (K'_{p-1}) but not to the subset (K''_{p-1}), let K_{j+1} denote the admissible rectangle[12] of cross-section K_j [§ 4, under Corollary 3], of norm d_{k+1}, and of first type, extending in the y_λ-direction. Then K_{j+1} has on its boundary the $(i+1)$-rectangles constructed, under (2), on K_i as base.

(3) Let all such rectangles as K_{j+1} be constructed.[13]

---

[12] These rectangles are all of diameter less than both $\frac{d_k}{4}$ and δ, since K_i (or K_j) is of diameter less than both $\frac{d_k}{32n}$ and $\frac{\delta}{8n}$. [§ 4, especially Corollaries 2 and 3].

[13] In case some of these rectangles of first type are partly outside $\bar{K}_p$, we shall retain only the parts which are inside. This case does not arise when $\bar{K}_p$ coincides with a subset of the rectangles (R^k).

(4) The portion of $\bar{K}_p$ outside the rectangles constructed under (1), (2), and (3) is made up of rectangles of the set $K^{\lambda}(d_{k+1})$[13] and rectangles of 0th type, norm d_{k+1} and edge-length d_{k+1} [see § 4].

Steps (1)–(4) complete a subdivision which satisfies the lemma. This may be verified directly with the aid of § 4, especially Results (I)–(III).

COROLLARY. *Suppose, in the above lemma, that $\bar{K}_p$ meets B only in points where $a_{k+1} \geqq \frac{1}{2\sqrt{n}}$. Then the lemma holds, even if* (B) *is replaced by the following:*

(B′) *A subset (K_p'') of (K_p') subdivides the rectangles of the set (R^k) on $\bar{K}_p$ in such a way that:*

(1) *The rectangles (K_p'') plus the rectangles of (R^k) which are not subdivided constitute a set of related rectangles of norm d_{k+1}, each of type not greater than the $(k+1)$th and each extending in the y_λ-direction for some $\lambda \leqq k+1$.*

(2) *Every rectangle of (K_p'') which meets B and has an edge parallel to the y_{k+1}-axis extends in the y_{k+1}-direction.*

(3) *In the set (K_p''), any two $(i+1)$-rectangles of type higher than the first, both of which extend in the y_{k+1}-direction, are either constructed on the same base or else on two bases having no common point.*

To establish this corollary, one need only go through the proof of the lemma, replacing λ by $k+1$.

7. **The rectangles (R^{k+1}). The subdivision theorem.** The following three steps lead from the set (R^k) of § 5 to the desired set, (R^{k+1}):

(I) An n-rectangle of the set (R^k) will be called an *outer rectangle* if its boundary contains any boundary point of the portion of B outside the rectangles (R^k). Let (K^k) denote the subset of (R^k) consisting of (1) the outer n-rectangles which meet B only in points where $a_{k+1} \geqq \frac{1}{2\sqrt{n}}$ and (2) the i-rectangles $(i = 0, \cdots, n-1)$ of (R^k) on the boundaries of these outer n-rectangles.

By successive applications of the corollary in § 6, let (K^k) be subdivided into rectangles of a new norm, d_{k+1}, of types from the 0th through the $(k+1)$th so that

(A) The rectangles of the subdivision plus the rectangles of the set (R^k) which are not subdivided form a set, (R^1), of related rectangles of norm d_{k+1}.

(B) Every rectangle of the subdivision which meets B and has an edge parallel to the y_{k+1}-axis extends in the y_{k+1}-direction.

(C) Every rectangle of the subdivision is of diameter less than $d_k/32n$.

(D) If two rectangles of type higher than the first in the subdivision extend in the y_{k+1}-direction and have any point common to their bases, then they have exactly the same base.

(II) The portion of B outside the rectangles (R^{I}) is bounded by a complex,[14] C_{n-2}, whose cells are cut from B by a subset of (R^{I}). Let C'_{n-2} denote the subcomplex of C_{n-2} cut out by the rectangles of the subdivision in Step I. Consider the subset (K^{I}) of (R^{I}) consisting of (1) those $(n-1)$-rectangles normal to the y_{k+1}-axis which cut out cells of C'_{n-2} at distance d_k or more from $(C_{n-2}-C'_{n-2})$, and (2) the rectangles of the set (R^k) on the boundaries of these $(n-1)$-rectangles. Every i-rectangle $(i=1,\cdots,n-1)$ of the set (K^{I}) which meets B is base of at most one admissible $(i+1)$-rectangle [§ 4, under Corollary 3], of norm d_{k+1}, lying outside the rectangles (R^{I}) and extending in the y_{k+1}-direction. Let each such rectangle be added to (R^{I}), along with the rectangles of the set $K^{k+1}(d_{k+1})$ [§ 4, Corollary 1] on its boundary and the rectangles of 0th type (§ 2), norm d_{k+1}, and edge-length d_{k+1} on the remainder of its boundary [§ 4 (III)].

The resulting set, (R^{II}), is a set of related rectangles.[15]

(III) The part of B outside the rectangles (R^{II}) is bounded on B by a complex,[14] Γ_{n-2}, whose cells are cut from B by a subset of (R^{II}). Let Γ'_{n-2} be the subcomplex of Γ_{n-2} cut out by the rectangles of the subdivision under Step I and the rectangles added under Step II. Consider any rectangle, K_p, which has no edge parallel to the y_{k+1}-direction and cuts out a cell of Γ'_{n-2} at distance d_k or more from the cells $(\Gamma_{n-2}-\Gamma'_{n-2})$. By Steps I and II, K_p is the base of two rectangles (K^1_{p+1}, K^2_{p+1}) of the set (R^{II}) which extend in the y_{k+1}-direction and which cut from B cells of the complex Γ'_{n-2}. The sum $(K^1_{p+1}+K_p+K^2_{p+1})$ is a rectangle, K_{p+1}, of first type with cross-section K_p (§ 4, Corollaries 2, 3). Let K_{p+1} replace the rectangles of which it is composed. Then K_{p+1} may contain points on the boundaries of two distinct rectangles, (K^1_{q+1}, K^2_{q+1}), $q>p$, of the set (R^{II}), where K^1_{q+1} and K^2_{q+1} both extend in the y_{k+1}-direction and have the same base, K_q. If so, let K^1_{q+1}, K_q, and K^2_{q+1} be amalgamated to form a rectangle, K_{q+1}, of first type and norm d_{k+1} with cross-section K_q. Then K_{q+1} has the whole of K_{p+1} on its boundary. Let such amalgamations be carried out until K_{p+1} is entirely on the boundary of every rectangle of higher dimensionality with which it is incident. Then let K_{p+1} be subdivided [see § 4 (II)] into rectangles of the set $K^{k+1}(d_{k+1})$ plus a set of rectangles of 0th type, norm d_{k+1}, and edge-length d_{k+1}.

When the above process is complete, we have a set (R^{III}) of related

[14] A consequence of the definition of related rectangles (§ 2).

[15] These statements are direct consequences of the construction and definitions. The results of § 4 are useful in verifying them. In Step II, the added rectangles are all of diameter less than $d_k/4$ and hence have no incidence relations with the rectangles of the set (R^k) which were not subdivided in Step I.

rectangles, containing the same points as (R^{II}). If, now, we add to (R^{III}) the subset of $K^{k+1}(d_{k+1})$ [§ 4, Corollary 1] consisting of the n-rectangles, with their boundaries, outside (R^{III}) and at distance not less than d_k from the cells $(\Gamma_{n-2} - \Gamma'_{n-2})$, then the newly added rectangles will be incident only with other rectangles of the set $K^{k+1}(d_{k+1})$.

The set, (R^{k+1}), which results from this last step, satisfies the Hypothesis (§ 5), *read with $(k+1)$ in place of k.*[15]

The above inductive process, carried out for $k = 1, \cdots, n-1$, leads to a set (R^n) of related rectangles of norm d_n and types from the 0th through the nth containing all points on B where one or more of the following inequalities holds:

$$a_i \geqq \frac{1}{\sqrt{n}} \qquad (i = 1, \cdots, n).$$

But since, by definition of the a's (§ 3),

$$\sum_{i=1}^{n} a_i^2 = 1,$$

(R^n) contains the whole of B. Let the rectangles of 0th type in the set (R^n), and also the part of the region R outside the rectangles (R^n) be divided into cubes of norm d_n and edge-length d_n. We then have a set of related rectangles (§ 2) containing $(R+B)$. This establishes the following theorem:

SUBDIVISION THEOREM. *If a region of euclidean n-space is bounded by a finite number of distinct regular $(n-1)$-topological $(n-1)$-manifolds* (§ 1), *then the region plus its boundary coincides with an n-complex. Each cell of the complex is the portion of the region or its boundary inside one of a finite set of distinct i-rectangles $(i = 0, \cdots, n)$. In the terminology of this paper, the i-rectangles are all of the same norm* (§ 4) *and are of the 0th through the nth types.*

LEHIGH UNIVERSITY, BETHLEHEM, PA.

LOCALLY HOMOGENEOUS SPACES IN DIFFERENTIAL GEOMETRY.[1]

By J. H. C. Whitehead.

1. **Introductory.** This note is about spaces which are locally equivalent to a space of some Lie group, in the sense of Klein's Erlanger Programm. They have been called *Clifford forms*[2] of the space of the Lie group by analogy with the spaces considered in the Clifford-Klein[3] problem. The Clifford-Klein problem is to characterize topologically all the Riemannian spaces (as originally stated by Klein, all the closed, or compact spaces) with constant curvature—that is, all the spaces which are locally hyperbolic, Euclidean, or elliptic. What may be called the fundamental lemma is due to H. Hopf[4], and states that, subject to a completeness condition, the general covering manifold of any locally Euclidean (hyperbolic, elliptic) space is an ordinary Euclidean (hyperbolic, spherical) space. In other words, if two spaces in the class considered are locally equivalent, and if they are both simply connected[5], then they are equivalent. We propose to prove this last assertion for the larger class of spaces defined below.

The spaces in question are n-dimensional manifolds of class[6] 2 with additional structure defined by a locally transitive infinitesimal group of Lie. That is to say, a locally transitive group of infinitesimal transformations is defined near each point, the infinitesimal groups at any two points being similar. This means that they are not only simply isomorphic, but that there is a point transformation of some n-cell containing one of the points into an n-cell containing the other, and the infinitesimal transformations at the former into linear combinations of the infinitesimal transformations at the latter.

In the terminology of the Cambridge tract referred to above, such a space

[1] Received March 23, 1932.

[2] E. Cartan, Annales de la Soc. Polonaise de Math., 8 (1929), p. 208.

[3] See F. Klein, Zur Nicht-Euclidischen Geometrie, Gesammelte Abhandlungen, vol. 1, pp. 353–383.

[4] Math. Annalen, 95 (1926), pp. 313–339. See also E. Cartan, Leçons sur la géométrie des espaces de Riemann, Paris, 1928, Chap. III.

[5] We shall use the term "simply connected" to mean that any closed curve can be deformed into a point. Nothing is implied about cycles of higher dimensionality.

[6] See Chap. VI of a recent Cambridge tract by O. Veblen and the present author called "The foundations of differential geometry", or an article in the Proc. Nat. Academy of Sciences, 17 (1931), pp. 551–561.

may be defined as an *n-dimensional manifold of class 2, whose pseudo-group*[7] *of automorphisms is transitive and is an r-parameter pseudo-group of Lie.*

These may be called *locally homogeneous spaces with a Lie pseudo-group,* and contain as a sub-class the homogeneous spaces with a Lie group[8].

Two locally homogeneous spaces are locally equivalent if, and only if, their infinitesimal groups are similar. Thus there are the "locally flat" affine, projective, and conformal spaces. In other cases the pseudo-group may be given indirectly, e. g. as the set of motions of a Riemannian space, or the collineations of a space of paths. Particularly interesting are the symmetric Riemannian spaces[9]—so called because the reflexion about any point preserves the Riemannian metric. These spaces are fundamental in Cartan's investigations of homogeneous spaces with simple groups[10]. Symmetric affine spaces of paths have been similarly defined[11], and their importance will doubtless increase with our knowledge about homogeneous spaces in general.

By an $\mathfrak{H}_n$ we shall mean a locally homogeneous space of n dimensions which is connected and which satisfies the following completeness condition.[12]

If U_n is a region in H_n, any particular $\mathfrak{H}_n$, and if the general covering manifold of U_n is an $\mathfrak{H}_n$, then $U_n = H_n$.

Our theorem is:

If two $\mathfrak{H}_n$'s are locally equivalent, and if both are simply connected, then they are equivalent.

It appears that the group of a simply connected $\mathfrak{H}_n$ is a Lie group

[7] A pseudo-group is a set of transformations operating on sub-sets of a space, any one of which may or may not include the whole space. In general the resultant of two transformations does not exist. If it does, it is in the set, and so is the inverse of each transformation. By a pseudo-group of Lie we mean one whose transformations are generated by an infinitesimal transformation group of Lie.

[8] Homogeneous spaces with a Lie group have been systematically studied by Cartan in a number of recent articles, a summary of which may be found in his Mémorial, La théorie des groupes finis et continus et l'analysis situs (Paris, 1930). This Mémorial will be referred to as G. C.

[9] Cartan, Bull. Soc. Math., 54 (1926), pp. 214–264, and 55 (1927), pp. 114–134.

[10] See, for instance, Cartan, Journal de Math., (9) 8 (1929), pp. 1–33.

[11] Cartan, Journal de Math., (9) 6 (1927), pp. 1–119, in particular pp. 80–90. An introductory account of symmetric affine spaces, starting with a somewhat different point of view, has been given by the present author (Math. Zeit. 35 (1932), pp. 644–659).

[12] Cf. H. Hopf and W. Rinow, Commentarii Math. Helvetici, 3 (1931), pp. 209–225. The condition "an $\mathfrak{H}_n$ is not equivalent to a proper sub-set of any $\mathfrak{H}_n$", which defines a class of spaces analogous to their $\mathfrak{F}_0$, is too weak for our purposes. It is satisfied, for example, by the covering surface of a Euclidean plane with a point removed. Topologically this surface is a 2-cell, but it is characterized by its metrical structure near the branch point, which is the only "missing" point at a finite distance from a given point. Cf. Cartan, Géométrie des espaces de Riemann, loc. cit., pp. 78–80.

which is locally isomorphic with its pseudo-group,[13] and also that any $\mathfrak{H}_n$ is analytic.

2. **Elementary deformations.** Let H_n be any $\mathfrak{H}_n$ and let g_r be its pseudo-group. Let P_0 be any point in H_n and x a coördinate system in which points near P_0 are represented by the box

$$|x^i| < 1,$$

the point P_0 corresponding to the origin. Transformations of g_r which operate on points near P_0 are given by equations of the form

$$\bar{x}^i = f^i(x, t^1, \cdots, t^r), \tag{2.1}$$

where the functions f are of class 2 in some region in the t space, and in some box

$$|x^i| < \delta < 1. \tag{2.2}$$

Moreover δ may be so chosen that some n-cell containing one of two given points in (2.2) is deformable into a cell containing the other by a continuous variation of the parameters t in (2.1). Such a transformation will be called an *elementary deformation* of the first point into the second. A cell such as (2.2), in which one of two given points can be deformed into the other by continuous variation of the t's in equations of the form (2.1), will be called an *elementary n-cell.*

Intuitively it is fairly obvious what is meant by deforming a point P_0 into a point P along a curve P_0P. We divide the curve into a number of segments P_0, P_1, P_2, $\cdots$, P_n, P, each of which is contained in an elementary n-cell. A deformation of P_0 into P along P_0P is the resultant of successive elementary deformations of P_0 into P_1, P_1 into P_2, $\cdots$, P_n into P.

We shall now prove the lemma. *A homogeneous space, S_n, with a Lie group, G_r, satisfies our completeness condition for an $\mathfrak{H}_n$.*

Let S_n cover a region U_n in a locally homogeneous space, h_n, and if possible let $U_n \neq h_n$. Let p^* be a point on the boundary of U_n. Since h_n is locally homogeneous, and since U_n is locally equivalent to S_n, h_n is locally equivalent to S_n. Therefore the pseudo-group of h_n is locally similar to G_r, and there is an elementary n-cell, C, containing p^*, such that to each elementary deformation in C corresponds a unique transformation of G_r. Since p^* is on the boundary of U_n there is a point, p_0, in C which is also in U_n. Let c be a curve joining p_0 to p^* without leaving C.

[13] The example of a locally Euclidean Möbius band, whose group of automorphisms has only one parameter, shows that this is not generally true.

There is no loss of generality in assuming p^* to be the only point on c which is not in U_n. For otherwise we could replace p^* by the first point on the boundary of U_n which a particle would meet in travelling along c from p_0 to p^*. Let P_0 be any point in S_n which covers p_0. Let t be a parameter for c which increases continuously from 0 to t^* as the corresponding point, $p(t)$, moves from p_0 to p^*. To the curve c and the point $p(t)$, $0 \leqq t < t^*$, corresponds a unique point, $P(t)$, in S_n, such that $P(0) = P_0$ and $P = P(t)$ is a continuous curve. To get our contradiction it will be sufficient to prove that $P(t)$ converges to a point, P^*, as $t \to t^*$.

Let $p(t)$ be continuously deformed along c into the point p^*, and let $a(t)$ be the elementary deformation of p_0 into $p(t)$. By our choice of the elementary n-cell C, there is a unique transformation, $A(t)$, of G_r, which corresponds to $a(t)$, and $A(t) \to A(t^*)$ as $t \to t^*$. Moreover $P(t) = A(t) P_0$ for $0 \leqq t < t^*$. Therefore $P(t) \to P^* = A(t^*) P_0$ as $t \to t^*$, and we have a contradiction in supposing S_n not to be complete.[14]

Any locally homogeneous space with a Lie pseudo-group is an analytic manifold. For among the allowable coördinate systems in which a given elementary n-cell is represented there is a uniquely determined set of analytically related coördinate systems, namely those for which the functions f in (2.1) are analytic[15] in x. These coördinate systems for any particular elementary n-cell, and for all the elementary n-cells contained in it, obviously satisfy the first group, A, of the axioms upon which our definition of an analytic manifold is based.[16] But the axioms C are satisfied by the set of all elementary n-cells, and it follows that the space is analytic.

From now on we confine our attention to analytic parameterizations, both for H_n and for g_r.

3. Let H_n be a simply connected $\mathfrak{H}_n$ with a pseudo-group g_r. The elementary deformations in a given elementary n-cell are generated by

[14] A similar argument can be used to show that our completeness axiom is equivalent to various other completeness conditions. For example an affine space of paths is com lete if, and only if, each path $P = P(s)$ is defined for $-\infty < s < \infty$, where s is an affine parameter (Hopf, loc. cit.). A Riemannian space is complete, or normal (Cartan), if it is locally compact with respect to the Riemannian metric. If the spaces in question are locally homogeneous both conditions are equivalent to ours. Cf. Hopf and Rinow, loc. cit.

[15] F. Schur, Math. Annalen, 41 (1893), p. 535. There is also a set of analytically related parameterizations, t, referred to which f are analytic in t. I take this opportunity of remarking that § 1 of Schur's paper contains a proof of the theorem given in § 10 of the paper by O. Veblen and myself (loc. cit.).

[16] Veblen and Whitehead, loc. cit. The axioms referred to are in three groups, A, B and C. We recall that A and C are sufficient to characterize the space, while B determines which coördinate systems, obtained according to certain rules from a fundamental set, are to be admitted as "allowable".

r vectors $\xi_1, \cdots, \xi_r$ $(r \geqq n)$, which are arbitrary to the extent of linear transformations with constant coefficients. The vectors ξ may be taken to define a local "parallelism", and our first step is to extend this parallelism into the large, thus defining a basis for the elementary deformations all over H_n. The parallelism in question is given in coördinates x by the completely integrable set of differential equations obtained from the equations

$$\xi^i(x) = c^\alpha \xi^i_\alpha(x) \qquad (\alpha = 1, \cdots, r), \tag{3.1}$$

by repeated differentiation and elimination of the constants[17] c. Writing $Z^1, \cdots, Z^m$ for $\xi^1, \cdots, \xi^n, \frac{\partial \xi^1}{\partial x^1}, \cdots, \frac{\partial^s \xi^n}{(\partial x^n)^s}$, these equations may be replaced by a first order set[18]

$$\begin{cases} \text{(a)} & \dfrac{\partial Z^A}{\partial x^i} + Z^B \Gamma^A_{Bi}(x) = 0 \qquad (A, B = 1, \cdots, m), \\ \text{(b)} & V^\varrho_A(x) Z^A = 0 \qquad (\varrho = 1, \cdots, m-r). \end{cases} \tag{3.2}$$

The system of functions[19] (Γ, V) is another set of components for the fundamental invariant of H_n, namely the infinitesimal transformations of g_r.

The differential equations (3.2) define a linear displacement[20] which is locally holonomic because (3.2) are completely integrable. Since H_n is simply connected this displacement is holonomic. There are, therefore, linearly independent solutions of (3.2), $Z_1(P), \cdots, Z_r(P)$, which are defined all over H_n, and which are uniquely determined by given initial values, $Z_\alpha(P_0) = A_\alpha$, chosen so as to satisfy (3.2b). The first n components of these solutions will be the components of the required vectorfields $\xi_1(P), \cdots, \xi_n(P)$.

All the elementary deformations may now be referred to a single set of canonical parameters.

Obviously

$$[\xi_\alpha, \xi_\beta] = C^\sigma_{\alpha\beta} \xi_\sigma,$$

with the same constants, C, near each point of H_n.

[17] See L. Bianchi, Lezioni sulla teoria dei gruppi continui finiti dei trazformazioni, Bologna, 1928, chap. IV, § 47.

[18] In case $r = n$, g_r then being simply transitive, these equations are

$$\frac{\partial \xi^i}{\partial x^k} + \xi^j \Gamma^i_{jk} = 0,$$

Γ being an affine connection. (Cf. L. P. Eisenhart, Proc. Nat. Academy of Sciences, 11 (1925), pp. 246–250.)

[19] Individually the sets of equations (3.2a) and (3.2b) will, in general, have no invariant significance. They will each depend on the equations chosen to solve for the constants c.

[20] The argument necessary for the passage from the small to the large is given in Veblen and Whitehead (Cambridge Tract, loc. cit., Chap. VII § 6).

4. By the third fundamental theorem[21] of Lie there is a group G_r, having $C^{\alpha}_{\beta\gamma}$ for its constants of structure. Let a basis having these constants be chosen for G_r. Then to each elementary deformation in H_n corresponds a unique element of G_r. Let O be a given point of H_n, and let G_{r-n} be the sub-group generated by elements corresponding to elementary deformations which leave O unaltered.

Let $\bar{H}_n$ be the projection of the group-manifold G_r by means of the co-sets[22] $t\,G_{r-n}$. Then $\bar{H}_n$ is a homogeneous space which is locally equivalent to H_n. Moreover $\bar{H}_n$ is an $\mathfrak{H}_n$ by the lemma of § 2. I say that H_n is equivalent to the general covering manifold of $\bar{H}_n$. This follows by an argument similar to the one used by O. Schreier[23] in proving the similar theorem for continuous groups. To a deformation of O along a curve OP to a variable point P corresponds a unique element, T, of G_r, namely the product of the elements corresponding to the elementary deformations composing the deformation $O \to P$. To P, therefore, corresponds a co-set, TG_{r-n}, and a point, $\bar{P}$, of $\bar{H}_n$. The point $\bar{P}$ is independent of the particular deformation $O \to P$. For let $O \to P$ be the resultant of elementary deformations $t_1, \cdots, t_m$, so that

$$P = t_m \cdots t_1 O.$$

The general deformation $O \to P$ along OP is of the form

$$P = t_m r_{m-1} t_{m-1} r_{m-2} \cdots t_1 r_0 O,$$

where r_α is a deformation of the point $P_\alpha = t_\alpha \cdots t_1 O$ into itself. But

$$r_\alpha = t_\alpha r'_{\alpha-1} t_\alpha^{-1},$$

where $r'_{\alpha-1}$ leaves $P_{\alpha-1}$ unaltered, and it follows that

$$t_m r_{m-1} t_{m-1} r_{m-2} \cdots t_1 r_0 O = t_m t_{m-1} r''_{m-2} \cdots t_1 r_0 O = t_m t_{m-1} \cdots t_1 R_0 O.$$

The transformation R_0 is the resultant of a finite number of elementary deformations which correspond to elements of G_{r-n}. Therefore R_0 corresponds to an element of G_{r-n}, and any deformation, $O \to P$, along OP

[21] This theorem was first proved in its complete form by Cartan, G. C., pp. 17–19, and Comptes Rendus, 190 (1930), pp. 914–916 and 1005–1007. Previous arguments had proved the existence of an element of a group (Gruppenkeim) having given constants of structure.

[22] Cf. Cartan, G. C., p. 25. The points of H_n are the co-sets of G_{r-n}, and an allowable coördinate-system for H_n is defined by projection on an analytic n-cell which meets no co-set in more than one point. From the special nature of G_{r-n} it is easy to see that such n-cells exist.

[23] Abhandlungen math. Seminar, Hamburg, 5 (1927), pp. 15–32.

corresponds to an element in the co-set TG_{r-n}. Therefore the point P, associated with a curve OP, corresponds to a unique point $\bar{P}$ in $\bar{H}_n$. The transformation $P \to \bar{P}$ is obviously unaltered by deformations of the curve OP, and is locally isomorphic. That is, it carries corresponding points under elementary deformations into corresponding points under G_r. Since H_n is simply connected it is independent of the curve OP, and H_n is equivalent to the general covering manifold of some region, V_n, in $\bar{H}_n$. By the completeness condition $V_n = \bar{H}_n$, and therefore H_n is equivalent to the general covering manifold of $\bar{H}_n$.

Let H_n^* be any simply connected $\mathfrak{H}_n$ which is locally equivalent to H_n, and let O^* be any point in H_n^*. Then there is an isomorphic point transformation of O into O^*, and points near the former into points near the latter. This transformation carries $\xi_1, \cdots, \xi_r$ into a set of vectors $\xi_1^*, \cdots, \xi_r^*$, which determine a basis for the elementary deformations in H_n^*. The basis ξ^* has the same constants of structure as ξ, and the deformations of O and O^* into themselves are generated by the same linear combinations of $\xi_1, \cdots, \xi_r$ and $\xi_1^*, \cdots, \xi_r^*$, respectively. It follows by the argument used above that H_n^* is equivalent to the general covering manifold of $\bar{H}_n$, and is therefore equivalent to H_n.

We have thus proved our theorem, which may also be stated as follows:

If two $\mathfrak{H}_n$'s are locally equivalent their general covering manifolds are equivalent.

THE COLORING OF GRAPHS.[1]

By Hassler Whitney.[2]

Introduction. In another paper, L,[3] the author has given a proof of a formula for $M(\lambda)$, the number of ways of coloring a graph in λ colors, due to Birkhoff. The numbers m_{ij}, in terms of which $M(\lambda)$ is expressed, are here studied in detail; a method of calculating them is given.

In Part I elementary properties of the m_{ij} and the m_i and of $M(\lambda)$ are given. The expression for the m_{ij} in terms of the broken circuits of the graph gives rise to sets of numbers α_i and β_i, in terms of which $M(\lambda)$ may also be expressed. In Part II it is shown that m_{ij} may be expressed as a polynomial in numbers $N_1, N_2, \cdots$, numbers of *non-separable* subgraphs, which are easier to count in a given graph than are the m_{ij}. Part III is devoted to the theory of an algebraic transformation. This is made use of in Part IV to show that if we know the linear terms f_{ij} of the above polynomial, the whole polynomial may be calculated. Finally, the polynomials for the f_{ij}, m_{ij} and m_i are calculated for a few values of the subscripts.

---

[1] Received February 16, 1932.—Presented to the American Mathematical Society, Oct. 25, 1930. This is, in revised form, the author's Harvard thesis. The section on recursion formulas has been left out; Part V has been added. An outline of the paper will be found in the Proceedings of the National Academy of Sciences, 17 (1931), pp. 122–125.

[2] National Research Fellow.

[3] References will be made to the following papers by the author.

I. A logical expansion in mathematics (L); Bulletin of the American Mathematical Society, 38 (1932), pp. 572–579.

II. Non-separable and planar graphs (N); Transactions of the American Mathematical Society, 34 (1932), pp. 339–362.

III. Characteristic functions and the algebra of logic (C. F.); to appear in the Ann. of Math.

IV. Congruent graphs and the connectivity of graphs (C. G.), American Journal of Mathematics, 54 (1932), pp. 150–168.

The notations in N will be made use of. We recall the following. The *rank* R and *nullity* N of a graph G containing V vertices and E arcs which is in P connected pieces are given by the equations $R = V - P$, $N = E - R = E - V + P$. A *subgraph* H of G is determined by naming a subset of the arcs of G. A graph is *non-separable* if it is connected and cannot be broken at a single vertex into two pieces. The *components* of a graph are the non-separable parts of the graph. The *sum* of two graphs G_1 and G_2 is the graph $G_1 + G_2$ containing the arcs and vertices of both graphs, provided they had no vertices in common. If they have common vertices, we consider these vertices as distinct in $G_1 + G_2$.

A graph is *colored* by assigning to each vertex a color in such a way that no two vertices which are joined by an arc are of the same color.

With reference to the four color map problem, we mention the following theorem. *If G is a planar graph and G' is a dual of G, and m_{ij} and m'_{ij} are their numbers, then*

$$m'_{ij} = m_{R-j,N-i}.^4$$

This follows immediately from the definition of dual graphs in N. Consider the class of graphs G with numbers m_{ij} for which the above numbers m'_{ij} are also the numbers for a graph G'; this class then includes all planar graphs (see N, Theorem 29). Hence a proposition which implies the four color map theorem is the following: *For any graph G of the above class,* $\sum_{i,j}(-1)^{i+j} m_{ij} 4^{V-i} > 0$. (We must, however, exclude graphs for which $m_{01} \neq 0$; see § 3.) This proposition is stronger than the four color map theorem; for there are graphs in the above class which are not planar.[5]

I. The Polynomial $M(\lambda)$; the Numbers α_i and β_i.

1. **Some properties of the m_{ij}.** We showed in L that if there were m_{ij} subgraphs of G of rank i, nullity j, then the number $M(\lambda)$ of ways of coloring G is[6]

$$M(\lambda) = \sum_{i,j}(-1)^{i+j} m_{ij} \lambda^{V-i} = \sum_i m_i \lambda^{V-i}, \tag{1.1}$$

$$m_i = \sum_j (-1)^{i+j} m_{ij}. \tag{1.2}$$

If G_1 and G_2 are the two graphs mentioned at the end of L (T_{21} and T_{32} in Fig. 2), whose arcs are $\alpha(ab)$, $\beta(ac)$, $\gamma(bc)$ and $\alpha(ab)$, $\beta(ac)$, $\gamma(bc)$, $\delta(bd)$, $\varepsilon(cd)$ respectively, we find for their numbers m_{ij} those given in the following tables.

G_1

rank \ nullity	0	1
0	1	0
1	3	0
2	3	1

G_2

	0	1	2
0	1	0	0
1	5	0	0
2	10	2	0
3	8	5	1

From these tables we immediately find their numbers m_i as given at the end of L. For instance, in G_1, m_2, the coefficient of λ, is $m_{20} - m_{21} = 3 - 1 = 2$.

[4] The formula of Theorem 8 of the Proceedings paper should be corrected.

[5] An example was recently found (May 13) by R. M. Foster: G contains the arcs ab, ac, bc, ad, de, ec, bf, fg, gc, eg, dg, fe. The corresponding graph G' is this same graph.

[6] Throughout this paper, sums are carried out over all non-zero terms unless otherwise stated. We put always $\binom{p}{q} = 0$ if $q<0$ or $q>p$.

m_{00} is always 1, for there is always exactly one subgraph of any graph (even of the null graph, containing no arcs or vertices), of rank 0, nullity 0, namely, that containing no arcs. If the graph contains no 1-circuits, each arc is of rank 1, nullity 0, and each subgraph of rank 1, nullity 0 contains just one arc. Thus $m_{10} = E$. We shall always assume in the future that there are no 1-circuits in the graph, unless expressly mentioned. Then also of course $m_{01} = 0$. In any case, $m_{10} + m_{01} = E$. There is just one subgraph of G of rank R, nullity N, namely G itself, and $m_{RN} = 1$. Thus

(1.3) $\quad m_{00} = m_{RN} = 1, \qquad m_{10} = E \qquad (\text{or } m_{10} + m_{01} = E).$

Of course

(1.4) $\qquad m_{ij} = 0 \text{ if } i<0, \ j<0, \ i>R \text{ or } j>N.$

There is a fundamental set of relations between the m_{ij}:

(1.5) $$m_{i0} + m_{i-1,1} + \cdots + m_{0i} = \binom{E}{i};$$

for the sum on the left is the total number of subgraphs of G of i arcs. We shall see in Part IV that there is no other linear, or even polynomial, equality connecting the m_{ij}.

2. Some properties of the m_i. We can derive immediately a number of properties of the coefficients m_i of $M(\lambda)$ from the interpretation of $(-1)^i m_i$ in terms of the broken circuits of G given in L. Let us show that $(-1)^R m_R > 0$. We drop out all the arcs of G we can without disconnecting any vertices that were formerly connected. The resulting graph H is a forest[7] of rank R. Order the arcs of G in a definite manner, choosing the arcs not in H first. Every circuit in G contains an arc not in H, and therefore every broken circuit contains an arc not in H. Thus H is a subgraph of G of R arcs not containing all the arcs of any broken circuit, and hence $(-1)^R m_R > 0$. Taking any subgraph of H, we see that

(2.1) $$(-1)^i m_i > 0, \qquad i = 0, 1, \cdots, R.$$

Also

(2.2) $$m_i = 0, \qquad i<0 \text{ or } i>R$$

from eq. (1.4). Hence $M(\lambda)$ has no constant term, as it would be m_V; but $V>R$ always. If G is connected, $R = V-1$, and the coefficient of λ in $M(\lambda)$ is $m_R \neq 0$.

3. Some properties of $M(\lambda)$. As $M(\lambda)$ has no constant term, $M(0) = 0$. This is obvious, as no graph can be colored in no colors.

Let us find $M(1)$.

$$M(1) = \sum_i \sum_j (-1)^{i+j} m_{ij},$$

[7] A graph of nullity 0.

or, if we sum diagonally,

$$M(1) = \sum_i (-1)^i \sum_k m_{i-k,k} = \sum_i (-1)^i \binom{E}{i}$$

by eq. (1.5). Hence

$$M(1) = (1-1)^E = 0$$

if $E>0$. If $E = 0$, $M(1) = m_{00} = 1$. Thus a graph containing an arc cannot be colored in one color; if it contains no arc, there is one way of coloring it in one color.

Let us find $M(\lambda)$ for a circuit of E arcs. Every subgraph of the circuit is of nullity 0 except the whole graph, which is of nullity 1. Hence

$$m_{i0} = \binom{E}{i}, \qquad 0 \leqq i \leqq E-1, \qquad m_{E-1,1} = 1.$$

Therefore

$$\begin{aligned} M(\lambda) &= \sum_{i=0}^{E} (-1)^i \binom{E}{i} \lambda^{V-i} - (-1)^E \lambda^{V-E} + (-1)^E \lambda^{V-(E-1)} \\ &= (\lambda-1)^E + (-1)^E (\lambda-1), \end{aligned} \tag{3.1}$$

remembering that $V = E$. We can also prove this by noting that there is just one broken circuit, which contains $E-1$ arcs. In particular, $M(2) = 1+(-1)^E$, which is 2 if E is even and 0 if E is odd.

Suppose the graph G contained a 1-circuit aa. Then in any coloring of the graph, a must be colored in a different color from itself, which is not possible. Consequently $M(0) = M(1) = \cdots = 0$, and thus $M(\lambda) \equiv 0$. We can prove this from L, eq. (11), by pairing off terms, so that the two terms of a pair are the same except that one contains the factor A_{aa} while the other does not. The two terms of each pair cancel.

4. **The numbers α_i and β_i.** We shall illustrate these numbers by means of an example, using the graph G_2 of § 1. Its circuits are $\alpha\beta\gamma$, $\gamma\delta\varepsilon$, $\alpha\beta\delta\varepsilon$; its broken circuits are $\alpha\beta$, $\gamma\delta$, $\alpha\beta\delta$. $(-1)^i m_i$ is the number of subgraphs of G_2 of i arcs not containing both α and β and not containing both γ and δ. For under these conditions, it does not contain α, β and δ either. Let R_i be the set of subgraphs of G_2 of i arcs, let $n_i(\alpha)$ be the number which contain the arc α, etc. Then

$$(-1)^i m_i = n_i(\overline{\alpha\cdot\beta}\cdot\overline{\gamma\cdot\delta}).$$

Expanding this as in C. F. into the second normal form, we have

$$\begin{aligned} (-1)^i m_i &= n_i(1-\alpha\beta-\gamma\delta+\alpha\beta\gamma\delta) \\ &= n_i(1)-n_i(\alpha\beta)-n_i(\gamma\delta)+n_i(\alpha\beta\gamma\delta). \end{aligned} \tag{4.1}$$

Consider a typical term, such as $n_i(\alpha\beta)$. This is the number of subgraphs of i arcs which contain both α and β. To form such a subgraph, we must pick out the two arcs α and β; we have left then $i-2$ arcs which we can pick out from the remaining $5-2=3$ arcs of G_2 in $\binom{5-2}{i-2}$ ways. This then is the value of the term. The sign of the term is $(-1)^l$, where l is the number of broken circuits combined in forming the term; in the above case, 1. In general, if the term is formed from l broken circuits, and if it contains k distinct arcs, then the value of the term is $(-1)^l \binom{E-k}{i-k}$.

Definition. Let α_{kl} be the number of sets of l broken circuits which together contain k arcs. Put

$$\alpha_k = \sum_l (-1)^l \alpha_{kl}. \tag{4.2}$$

Then

$$(-1)^i m_i = \sum_{k,l} (-1)^l \binom{E-k}{i-k} \alpha_{kl} = \sum_k \binom{E-k}{i-k} \alpha_k.$$

We note that α_k is just the sum of all the coefficients in eq. (4.1) which contain k arcs, and thus α_k is the α_k defined in C. F. at the end of § 10.

If we expand eq. (4.1) into the third normal form, we find

$$(-1)^i m_i = n_i[\bar\alpha\bar\gamma + \bar\alpha\bar\delta + \bar\beta\bar\gamma + \bar\beta\bar\delta - (\bar\alpha\bar\gamma\bar\delta + \bar\alpha\bar\beta\bar\gamma + \bar\alpha\bar\beta\bar\delta + \bar\beta\bar\gamma\bar\delta) + \bar\alpha\bar\beta\bar\gamma\bar\delta]. \tag{4.3}$$

We define β_k as being the sum of the coefficients of all terms in this equation containing k arcs, again as in C. F.

With the help of C. F., eq.'s (22) and (23), we can express any of the sets m_i, α_i, β_i in terms of any other of the sets. The relations are[8]

[8] The formulas not already proved may be proved from the others by use of the relation

$$\sum_s (-1)^s \binom{p-s}{m} \binom{q}{s} = \binom{p-q}{m-q}, \quad \text{if } p \geqq q;$$

see Netto, Lehrbuch der Combinatorik, Leipzig, 1927, § 158, (27). For instance, to prove the formula for α_i in terms of m_k, put

$$(-1)^i a_i = \sum_k \binom{E-k}{i-k} m_k = \sum_k \binom{E-k}{i-k} (-1)^k \sum_l \binom{E-l}{k-l} \alpha_l$$
$$= \sum_l \alpha_l \sum_k (-1)^k \binom{E-k}{E-i} \binom{E-l}{k-l}.$$

Replacing k by $k+l$, we find

$$(-1)^i a_i = \sum_l (-1)^i \alpha_l \sum_k (-1)^k \binom{E-l-k}{E-i} \binom{E-l}{k}$$
$$= \sum_l (-1)^i \alpha_l \binom{0}{l-i} = (-1)^i \alpha_i.$$

Thus $a_i = \alpha_i$, as required.

$$(-1)^i m_i = \sum_k \binom{E-k}{i-k} \alpha_k = \sum_k \binom{E-k}{i} \beta_k,$$

$$(4.4) \qquad (-1)^i \alpha_i = \sum_k \binom{E-k}{i-k} m_k = \sum_k \binom{k}{i} \beta_k,$$

$$(-1)^i \beta_i = (-1)^E \sum_k \binom{k}{E-i} m_k = \sum_k \binom{k}{i} \alpha_k.$$

The numbers for G_2 are

$$\alpha_0 = 1, \quad \alpha_2 = -2, \quad \alpha_4 = 1;$$

$$\beta_2 = 4, \quad \beta_3 = -4, \quad \beta_4 = 1;$$

$$m_0 = 1, \quad m_1 = -5, \quad m_2 = 8, \quad m_3 = -4, \quad m_4 = 0.$$

The first non-zero β_i is $\beta_{E-R} = \beta_N$. Remembering that $m_k = 0$, $k > R$, we find from the last equation above

$$(4.5) \qquad \beta_N = (-1)^R m_R,$$

which is the last non-zero coefficient in $M(\lambda)$; the coefficient of λ, if the graph is connected.

In terms of the α_i, we have

$$(4.6) \qquad M(\lambda) = \lambda^{V-E} \sum_i (-1)^i \alpha_i (\lambda - 1)^{E-i}.$$

For, calling this expression $\varphi(\lambda)$,

$$\varphi(\lambda) = \lambda^{V-E} \sum_i (-1)^E \alpha_i (1-\lambda)^{E-i}$$

$$= \lambda^{V-E} \sum_i (-1)^E \alpha_i \sum_k (-1)^k \binom{E-i}{k} \lambda^k$$

by the binomial formula,

$$= \sum_k \lambda^{V-E+k} (-1)^{E-k} \sum_i \binom{E-i}{k} \alpha_i,$$

or, replacing k by $E-k$,

$$\varphi(\lambda) = \sum_k \lambda^{V-k} (-1)^k \sum_i \binom{E-i}{E-k} \alpha_i$$

$$= \sum_k m_k \lambda^{V-k} = M(\lambda).$$

We shall see by examples in Part V that the α_i are given by simpler formulas in terms of the numbers of non-separable subgraphs of G than are the m_i.

II. m_{ij} IN TERMS OF NON-SEPARABLE SUBGRAPHS.

5. **The numbers N_1, N_2, ...** To find m_{ij}, we count all the subgraphs of G of rank i, nullity j. In this part we shall show that it is

not necessary to count all the subgraphs; we need count only the non-separable subgraphs. We say a subgraph is non-separable if it becomes so when any isolated vertices there may be are dropped out. This operation alters neither the rank nor the nullity of the subgraph.

We note that Theorem 7A holds even if we allow the graphs to contain 1-circuits. If G' is G except for its m_{01} 1-circuits, then obviously

$$m_{ij} = \sum_k \binom{m_{01}}{k} m'_{i,j-k}. \tag{5.1}$$

Let us say two graphs are of the same type if they are congruent. Make a list of all types of non-separable graphs; call these types T_1, T_2, T_3, $\cdots$. There are of course but a finite number of types of graphs with a given number of arcs. Let N_i be the number of subgraphs of G of type T_i, $i = 1, 2, \cdots$. The fundamental theorem of this part is that m_{ij} for any graph G is given by a unique polynomial in the numbers N_1, N_2, $\cdots$ for G. When we speak of a polynomial, we mean always a polynomial in a finite number of variables. Thus we do not allow expressions such as $N_1 + N_2 + N_3 + \cdots$, even if, for any given graph, there are but a finite number of non-vanishing terms.

THEOREM 5A. *Any polynomial in the numbers N_1, N_2, $\cdots$, which equals 0 for all graphs is identically 0.*

Suppose the theorem false. Then there is a polynomial which is 0 for all graphs but not $\equiv 0$. Let P be such a polynomial containing the least possible number, say q, of variables. Let us call these variables N_1, N_2, $\cdots$, N_q, where T_1, corresponding to N_1, is one of the types of graphs concerned containing the least number of arcs.

Arrange the polynomial according to descending powers of N_1, and say it is of degree α in this variable:

$$P(N_1, N_2, \cdots, N_q) = P_0(N_2, \cdots, N_q) N_1^{\alpha} + P_1(N_2, \cdots, N_q) N_1^{\alpha-1} + \cdots + P_\alpha(N_2, \cdots, N_q), \; P_0 \not\equiv 0.$$

Take now any graph G, and form from it graphs $G_1, G_2, \cdots, G_\alpha$, each being formed from the last by adding a graph of type T_1, letting it have no vertices in common with the preceding graphs. As a graph of type T_1 contains no subgraph of any of the types $T_2, \cdots, T_q$, we have in this manner added no subgraphs of any of these types. Hence the number N_1 for each graph G_i is one greater than that for the preceding graph, G_{i-1}, while the numbers $N_2, \cdots, N_q$ are the same. P vanishes for these distinct values of N_1 and the coefficients $P_0, P_1, \cdots, P_\alpha$ are constant for these graphs, and hence these coefficients must vanish for these graphs. They vanish, in particular, for the graph G. But G was any graph, so they vanish for all graphs. But this is a contradiction. For, if q was > 1,

P_0 is a polynomial in a smaller number of variables, vanishing for all graphs without being identically 0, and if q was 1, P_0 is a constant $= 0$, and P was not of degree α in N_1.

THEOREM 5B. *Theorem* 5A *holds if the words "for all graphs" are replaced by the words "for all non-separable graphs".*

Using the same notation as before, let G be any non-separable graph. The only change necessary is to make $G_1, \cdots, G_\alpha$ non-separable also. Suppose first that the numbers $N_1, \cdots, N_q$ do not include E, the number of arcs in G. Let s be the greatest number of arcs in a graph of any of the types $T_1, \cdots, T_q$. We add a graph of type T_1 to G, and then join two of its vertices to two vertices of G by distinct chains, each containing at least $s/2$ arcs. The resulting graph G_1 is non-separable. Moreover, in adding the two chains we have added no subgraphs of any of the types $T_2, \cdots, T_q$. $G_2, \cdots, G_\alpha$ are formed similarly. If, however, the polynomial contains the variable E, we merely add a suspended chain of s arcs each time, increasing E by s, and leaving the other N_i unchanged.

6. m_{ij} **in terms of** $N_1, N_2, \cdots$. Given a graph G, consider the set S of all its subgraphs of rank i, nullity j. Drop out any isolated vertices there may be from each subgraph. Let us list all possible graphs of rank i, nullity j which contain no isolated vertices (many may be separable), and divide them into types $U_1, U_2, \cdots, U_p$. We say here two graphs are of the same type if they are equivalent.[8a] Each subgraph of S is of one of these types; say S_k, containing s_k members, is the set of subgraphs of type U_k (any or all of the s_k may be 0). Then

$$m_{ij} = s_1 + s_2 + \cdots + s_p. \tag{6.1}$$

We shall show how to evaluate each s_k.

Let H be a graph of type U_k, let $H_1, H_2, \cdots, H_q$ be its components, named in a definite order, (there will be but one component if H is non-separable), and say H_i is of type T_i, $i = 1, \cdots, q$. (Some of these types may be the same.) We now define a set R_k as follows: Any member of R_k consists of an ordered set of non-separable subgraphs of G. We name first in G a subgraph H_1' of type T_1, next a subgraph H_2' of type T_2 (which may have parts in common with H_1'), $\cdots$, and finally a subgraph H_q' of type T_q (which may have parts in common with any of the preceding graphs). Let n_k be the number of sets of graphs in R_k. If, for instance, H consists of two arcs, its components H_1 and H_2 each being an arc, R_k is the set of ordered pairs of arcs in G, the two arcs being distinct or not, and $n_k = E^2$.

[8a] Two graphs are equivalent if they have the same components.

Consider any member of R_k. It may happen that some of the subgraphs picked out of G to form this member, say the first three, form a circuit of graphs in G. If so, we shall say that this member of R_k has the property

$$H_1 \circ H_2 \circ H_3 .$$

We are interested in those members of R_k in which no subset of the subgraphs form a circuit of graphs in G. These are

$$R_k(\overline{H_1 \circ H_2} \cdot \overline{H_1 \circ H_3} \cdots \overline{H_{q-1} \circ H_q} \cdot \overline{H_1 \circ H_2 \circ H_3} \cdots \overline{H_1 \circ H_2 \circ \cdots \circ H_q})$$
$$= R_k(\bar{A}_{12} \cdot \bar{A}_{13} \cdots \bar{A}_{q-1,q} \cdot \bar{A}_{123} \cdots \bar{A}_{12\ldots q}) = R_k(A).$$

Remembering that the types $T_1, \cdots, T_q$ may not be distinct, let us pick out the distinct types $T_1', T_2', \cdots, T_r'$ from these. Say the first τ_1 types $T_1, \cdots, T_{\tau_1}$ are the same type as T_1', etc. Of course

$$\tau_1 + \tau_2 + \cdots + \tau_r = q. \tag{6.2}$$

We shall now show that

$$s_k = \frac{1}{\tau_1!\, \tau_2! \cdots \tau_r!}\, n_k(A). \tag{6.3}$$

To this end we establish a correspondence between the members of S_k and of $R_k(A)$. Consider first any member of $R_k(A)$; say $H_1', \cdots, H_q'$ are the subgraphs of G named, where H_i' is of type T_i. Let H' be the subgraph of G containing the vertices and arcs of all these graphs. No subset of these graphs form a circuit of graphs, and hence these graphs are the components of H', by N, Theorem 17. Therefore H' is of the type U_k, and is a member of S_k. Thus to each member of $R_k(A)$ corresponds just one member of S_k.

Consider now any member H' of S_k in G, whose components are $H_1', \cdots, H_q'$, H_i' being of type T_i. There are exactly $\tau_1!\, \tau_2! \cdots \tau_r!$ corresponding members of $R_k(A)$, obtained as follows. $H_1', \cdots, H_{\tau_1}'$ all being of the same type T_1', and these being distinct in G, we name first any one of these, then any other, etc., which we can do in $\tau_1!$ ways, forming the first part of an ordered set of graphs of $R_k(A)$. The next τ_2 subgraphs we can name in $\tau_2!$ ways, etc. The equation now follows. This is true in particular if all the types $T_1, \cdots, T_q$ are different or all the same; also if H' is non-separable, in which case, there is but one type $T_1 = T_1'$, and $\tau_1 = 1$.

Summing over k, we have

$$m_{ij} = \sum_k \frac{1}{\tau_1!\, \tau_2! \cdots \tau_r!}\, n_k(A). \tag{6.4}$$

This expression is evaluated by expanding A by the logical expansion:

$$m_{ij} = \sum_k \frac{1}{\tau_1!\,\tau_2!\cdots\tau_r!}\, n_k[1-(A_{12}+A_{13}+\cdots+A_{q-1,q} \\ +A_{123}+\cdots+A_{12\ldots q})+(A_{12}A_{13}+\cdots)-\cdots]. \tag{6.5}$$

7. **Evaluation of eq. (6.5).** We shall show that a typical term of the above expression is a polynomial in the numbers $N_1, N_2, \cdots$. This will prove that m_{ij} is a polynomial in these numbers.

Consider the term

$$n_k(A_{i_1 i_2 \ldots i_\alpha} A_{j_1 j_2 \ldots j_\beta} \cdots A_{l_1 l_2 \ldots l_\delta}) = n_k(B_1 B_2 \cdots B_u) = n_k(B), \tag{7.1}$$

each B_i stating that a certain subset of the graphs $H_1', \cdots, H_q'$ form a circuit of graphs in G. This term is the number of ways of naming in G first a subgraph H_1' of type $T_1, \cdots$, finally a subgraph H_q' of type T_q, so that the relations $B_1, \cdots, B_u$ all hold.

Divide the relations $B_1, \cdots, B_u$ into sets as follows. Suppose two of these included the same graph; e. g. if

$$B_1 = H_1 \circ H_2, \qquad B_2 = H_2 \circ H_4 \circ H_5,$$

then B_1 and B_2 both include the graph H_2. Say then that B_1 and B_2 are connected. If B_i and B_j, also B_j and B_k, are connected, say that B_i and B_k are connected. We put two relations into the same set if they are connected, otherwise, into different sets. Let C_1 stand for all the relations in the first set, C_2, for all those in the second set, $\cdots$, $C_{v'}$ for all those in the last set. Then

$$B = C_1 C_2 \cdots C_{v'}. \tag{7.2}$$

Divide now the graphs $H_1, H_2, \cdots, H_q$ into sets as follows: All the graphs appearing in any relation B_i of C_1 we put in the first set Q_1, all appearing in a relation of C_2 we put in the second set $Q_2, \cdots$, all appearing in the last set $C_{v'}$ we put in the set $Q_{v'}$. There may be some graphs left over (if we are considering the first term of eq. (6.5), there are no B's, and none of the graphs $H_1, \cdots, H_q$ have been placed in sets); if so, put each of these in a set by itself, calling these sets $Q_{v'+1}, \cdots, Q_v$. If we define the characteristic functions

$$C_i = 1, \qquad i = v'+1, \cdots, v, \tag{7.3}$$

then we may write also

$$B = C_1 C_2 \cdots C_v, \tag{7.4}$$

where $v = v'$ if there were no graphs left over. Now each graph H_i falls into exactly one of the sets $Q_1, \cdots, Q_v$.

In determining $n_k(B)$, we picked out the graphs $H_1', \cdots, H_q'$ from G in a definite order. Let us fix this order so that the first graphs fall into Q_1, the next fall into Q_2, etc. We can now say that $n_k(B)$ equals the number of ways of picking out from G first the graphs of Q_1 in a definite order so that C_1 holds, next the graphs of Q_2 in a definite order so that C_2 holds, $\cdots$, finally the graphs of Q_v in a definite order so that C_v holds. (For values of $i > v'$, the words "so that C_i holds" may of course be left out.) Hence if $n_k'(C_i)$ is the number of ways of picking out from G graphs of the set Q_i in a definite order so that C_i holds, then

$$n_k(B) = n_k'(C_1)\, n_k'(C_2) \cdots n_k'(C_v). \tag{7.5}$$

We shall show that each term of this product, such as $n_k'(C_j)$, is a homogeneous linear polynomial in the numbers $N_1, N_2, \cdots$, showing that $n_k(B)$ is a homogeneous polynomial of degree v in these numbers.

Let us say $H_1, H_2, \cdots, H_z$ are the graphs in Q_j. Name graphs of these types in G so that C_j holds. These subgraphs of G together form a subgraph I of G. I is non-separable. For if first C_j contains no B's, i. e. $C_j = 1$, then there is but a single graph H_1' in Q_j, and I is H_1', a non-separable graph. Otherwise, I consists of a number of these graphs, various ones of which form circuits of graphs. Suppose

$$C_j = B_{i_1} B_{i_2} \cdots B_{i_l}.$$

Consider that part of I, I_1, formed of the graphs forming the circuit of graphs corresponding to B_{i_1}. By N, Theorem 16, I_1 is non-separable. Similarly, $I_2, \cdots, I_l$ are non-separable. Now the relations $B_{i_1}, \cdots, B_{i_l}$ are connected. Hence there is a graph H_s' (containing at least two vertices) in both B_{i_1} and some other relation, say B_{i_2}, i. e. in both I_1 and I_2. Thus these two graphs form a circuit of graphs, and the two together form a non-separable graph. Continuing in this manner, we see finally that I is non-separable.

Fit the graphs $H_1, \cdots, H_z$ of Q_j together in all possible ways so that the relation C_j holds. We get various non-separable graphs, of types $T_1'', T_2'', \cdots, T_w''$ say. Suppose I is of type T_i''. Let η_i equal the number of ways of fitting first H_1, next H_2, $\cdots$, finally H_z into I so that C_j holds and I is completely filled out. Then if there are N_i subgraphs in G of type T_i'', $i = 1, \cdots, w$,

$$n_k'(C_j) = \eta_1 N_1 + \eta_2 N_2 + \cdots + \eta_w N_w. \tag{7.6}$$

To prove this, consider the correspondence between the sets of graphs counted in $n'_k(C_j)$ and the subgraphs of G of types $T''_1, \cdots, T''_w$. Corresponding to each way of picking out a set of graphs from G in counting $n'_k(C_j)$ is a non-separable subgraph I of G of one of these types. Consider any such subgraph I of type T''_i of G. There are exactly η_i ways of naming subgraphs in G properly so that these graphs together form I in G. Hence each subgraph of this type contributes an amount η_i to $n'_k(C_j)$; as there are N_i such subgraphs in G, they all contribute an amount $\eta_i N_i$, and the result is established.

As the numbers $\tau_1, \tau_2, \cdots, \tau_r$, and $\eta_1, \eta_2, \cdots, \eta_w$ depend only on the types of graphs considered and on i and j, not on the graph G, we have proved all but the statement of uniqueness in the

THEOREM 7A. *m_{ij} for any graph G is given by a unique polynomial in the numbers $N_1, N_2, \cdots$, containing no constant term, the coefficients being independent of the graph G.*

To prove the uniqueness of the polynomial, suppose there were two such polynomials. Then their difference is a polynomial which is 0 for all graphs, and hence is identically 0, by Theorem 5A; the two polynomials are therefore identically equal.

Examples. We consider here only graphs containing no 1- or 2-circuits.

m_{20}: Each subgraph H of G of rank 2, nullity 0, consists of two distinct arcs, H_1 and H_2. As H_1 and H_2 are of the same type, $\tau_1 = 2$. Hence, by eq. (6.5),

$$m_{20} = \frac{1}{2!} n(\overline{H_1 \circ H_2}) = \frac{1}{2} n(1 - H_1 \circ H_2).$$

$n(1)$ is the number of ways of picking out first any arc, then again any arc from G. This is E^2. $n(H_1 \circ H_2)$ is the number of ways of picking out first an arc H_1, then an arc H_2, such that the two form a circuit of graphs in G. As there are no 2-circuits in G, these arcs must be identical. This number is thus E. This gives the obvious result

$$m_{20} = \frac{1}{2}(E^2 - E) = \binom{E}{2}.$$

m_{31}: A subgraph H of G of rank 3, nullity 1 consists either of a quadrilateral, or of a triangle H_1 and an arc H_2. The number of subgraphs of the first type is N_{31} say. The number in the second set is, as $\tau_1 = \tau_2 = 1$, $n(1 - H_1 \circ H_2)$. Now $n(1) = E N_{21}$ if there are N_{21} triangles in G. If H_1 and H_2 are to form a circuit of graphs, the arc H_2 must be one of the arcs of the triangle H_1. Together they form a triangle I. H_1 can be fitted into I in one way; the arc can be fitted in in three ways. Hence $\eta_1 = 3$, and $n(H_1 \circ H_2) = 3 N_{31}$. Adding,

$$m_{31} = N_{31} + (E-3)\,N_{21},$$

a result which can be easily verified.

8. **The types of non-separable subgraphs contributing to a given m_{ij}.** Given two numbers i, j, we wish to know what numbers $N_1, N_2, \cdots$, appear in the polynomial for m_{ij}. We are interested especially in the linear terms of the polynomial; c. f. Part IV. If the number N_s corresponding to the type T_s of graph appears in some term of the expansion (6.5) for m_{ij}, we shall say a graph of this type *contributes* to m_{ij}. However, it might possibly happen that the terms cancel out so that N_s does not actually appear in the polynomial for m_{ij}.

As we saw in § 7, a term such as $n_k(B)$ in eq. (6.5) is linear if all the graphs $H_1, H_2, \cdots, H_q$ fall into a single set Q_1, i. e. if $v = 1$. This happens when the relations $B_1, \cdots, B_u$ are connected and each graph H_i falls into one of these relations, or if H is non-separable.

From the considerations of § 6 and § 7 we have the

THEOREM 8A. *The necessary and sufficient condition that a non-separable graph I contribute to the linear terms of m_{ij} is that there exists a graph H of rank i, nullity j, with the components $H_1, H_2, \cdots, H_q$ ($q \geqq 1$), such that if these graphs are allowed to coalesce in the proper manner, I is formed.*

As an immediate application, an arc contributes to m_{i0} only. It is seen that a non-separable graph can only contribute to m_{ij} if it contributes to the linear terms of some m_{kl}, $k \leqq i$, $l \leqq j$. (This follows also directly from the facts in Part IV.)

LEMMA A. *A non-separable graph of rank i, nullity j, contributes to m_{ij} with the coefficient* 1.

These are the only non-separable graphs that are counted directly in eq. (6.5).

LEMMA B. *A non-separable graph of rank $i-k$, nullity $j+k$, contributes to the linear terms of m_{ij} if $k>0$.*

Let I be such a graph. As it is of nullity >0, it contains at least two arcs. Pull out one of its arcs, and let I' be the new graph, composed of this arc, I_1', and the remaining part of I, I_2'. I' is of rank $i-k+1$, nullity $j+k-1$. Separate I_2' into its components if it is separable, which does not alter its rank or nullity. If I' is still of nullity $>j$, we take one of its components of nullity >0, and pull out an arc as before. Continuing, we obtain finally a set of non-separable graphs, the whole being of rank i, nullity j, which, if pieced together properly, form I. The lemma now follows from Theorem 8A.

LEMMA C. *A non-separable graph of rank $i-k-l$, nullity $j+k$, contributes to the linear terms of m_{ij} if $k \geqq 0$, $l>0$.*

Let I be such a graph. Add to it l arcs, forming the graph I' in $l+1$ connected pieces. I' is of rank $i-k$, nullity $j+k$. Pull out arcs from this graph as in Lemma B until we obtain a graph of rank i, nullity j. I is formed from this graph by reconstructing I', then letting the arcs coalesce with arcs of the resulting piece.

Various graphs of rank $i-k$, nullity $j-l$, may also contribute. If so, similar graphs of rank $i-k-s$, nullity $j-l$, contribute.

LEMMA D. *No non-separable graph containing more than $i+j$ arcs contributes.* Obviously, from Theorem 8A.

LEMMA E. *No non-separable graph of rank $i+l$, nullity $j-k$, contributes to m_{ij} if $l \geqq 0$, $k>0$.*

For suppose I were such a graph which contributes to m_{ij}. Then, by Theorem 8A, I is formed by letting the components $H_1, H_2, \cdots, H_q$, of which there are at least two, of a graph H of rank i, nullity j, coalesce. But then, by N, Theorem 14, I is of rank $<i$, a contradiction.

LEMMA F. *If m_{ij}, $j>0$, is >0 for some graph G, then there exists a non-separable graph I of rank i, nullity j (which contributes).*

G contains a subgraph H of rank i, nullity j. If it is separable, let $H_1, \cdots, H_q$ be its components. Let an arc of H_1 coalesce with an arc of H_2; this reduces the rank of the whole by one, but leaves the nullity unchanged. Continue in this manner until we have a non-separable graph H' of rank $i-q+1$, nullity j. Let ab be an arc of H'; replace it by two arcs, ac and cb, c being a new vertex. This increases the rank by one but leaves the nullity unchanged. Continuing, we have finally the non-separable graph I as required.

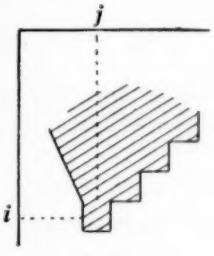

Fig. 1.
Showing those types of non-separable graphs contributing to the linear terms of a given m_{ij}

These results are collected in Fig. 1.

III. THE TRANSFORMATION T.

9. **Definition and elementary properties of T.** Let there be a set of elements $a, b, c, \cdots$, such that for any two elements a, b, there is a "sum element" $a+b=b+a$ also in the set. (We shall later interpret these elements as graphs.) Let $f_i(a)$, $i=1, 2, \cdots$, be a set of numbers defined for each element a, with the property that

$$f_i(a+b)=f_i(a)+f_i(b) \tag{9.1}$$

for any two elements a, b. We wish to find a transformation T from these sets of numbers to sets of numbers $m_i(a)$, $i=0, 1, 2, \cdots$, with the properties that $m_0 \equiv 1$ and

$$(9.2)\qquad \begin{aligned} m_i(a+b) &= m_0(a)\,m_i(b)+m_1(a)\,m_{i-1}(b)+\cdots+m_i(a)\,m_0(b)\\ &= \sum_k m_k(a)\,m_{i-k}(b).^{9}\end{aligned}$$

The following symbol will be found convenient: $R_i^k(f)$ is the sum of all terms formed by multiplying together k numbers $f_{j_1}, f_{j_2}, \cdots, f_{j_k}$, whose subscripts (which need not be distinct) add up to i:

$$(9.3)\qquad j_1+j_2+\cdots+j_k = i,$$

and two terms are called different if their factors are arranged in different orders. Put also

$$(9.4)\qquad R_0^0(f)=1,\quad R_i^0(f)=0,\quad i>0,\quad R_i^k(f)=0,\quad i<k.$$

Thus for example

$$(9.5)\qquad \begin{aligned} R_4^1(f) &= f_4,\\ R_4^2(f) &= f_1f_3+f_2f_2+f_3f_1 = 2f_1f_3+f_2^2,\\ R_4^3(f) &= f_1f_1f_2+f_1f_2f_1+f_2f_1f_1 = 3f_1^2f_2,\\ R_4^4(f) &= f_1f_1f_1f_1 = f_1^4.\end{aligned}$$

Now for any element a, let $f_1(a)$, $f_2(a)$, $\cdots$, be the coefficients of a power series in x:

$$F_a(x) = f_1(a)\,x+f_2(a)\,x^2+\cdots.$$

We shall define $m_i(a)$ as being the coefficient of x^i in the power series expansion of $e^{F_a(x)}$:

$$\begin{aligned} e^{F_a(x)} &= 1+F_a(x)+\frac{1}{2!}F_a^2(x)+\cdots\\ &= m_0(a)+m_1(a)x+m_2(a)\,x^2+\cdots = M_a(x).\end{aligned}$$

Equating coefficients determines the m_i. The first few are

$$(9.6)\qquad \begin{aligned} m_0 &= 1,\quad m_1 = f_1,\quad m_2 = f_2+\frac{1}{2}f_1^2,\\ m_3 &= f_3+\frac{1}{2!}(f_1f_2+f_2f_1)+\frac{1}{3!}f_1^3 = f_3+f_1f_2+\frac{1}{6}f_1^3,\\ m_4 &= f_4+f_1f_3+\frac{1}{2}f_2^2+\frac{1}{2}f_1^2f_2+\frac{1}{24}f_1^4,\end{aligned}$$

and the general term is found to be

$$m_i = \sum_k \frac{1}{k!}R_i^k(f).$$

[9] See footnote [6] in § 1, Part. I.

m_i is a polynomial in $f_1, f_2, \cdots, f_i$, for $i>0$. If we collect terms, we find the following expression for the transformation T:

$$(9.7) \qquad m_i = \sum_k \frac{1}{k!} R_i^k(f) = \sum_{\sum k\alpha_k = i} \frac{1}{\alpha_1!\,\alpha_2!\cdots\alpha_i!} f_1^{\alpha_1} f_2^{\alpha_2}\cdots f_i^{\alpha_i},$$

where in the second sum all terms with $\alpha_1+2\alpha_2+\cdots+i\alpha_i = i$ appear.

To find the inverse of T, we compare coefficients of x^i in the equation

$$F_a(x) = \log M_a(x).$$

We find

$$(9.8) \qquad f_i = \sum_k \frac{(-1)^{k-1}}{k} R_i^k(m) = \sum_{\sum k\alpha_k = i} \frac{(-1)^s s!}{\alpha_1!\,\alpha_2!\cdots\alpha_i!} m_1^{\alpha_1} m_2^{\alpha_2}\cdots m_i^{\alpha_i},$$

where

$$s = \alpha_1+\alpha_2+\cdots+\alpha_i-1.$$

The first few f_i are

$$(9.9) \qquad f_1 = m_1, \qquad f_2 = m_2-\frac{1}{2}m_1^2, \qquad f_3 = m_3-m_1m_2+\frac{1}{3}m_1^3.$$

If we write T in the form

$$(9.10) \qquad m_i = f_i+\sum_{k\geq 2} \frac{1}{k!} R_i^k(f),$$

each term of the sum contains $k \geqq 2$ factors. Similarly for T^{-1}. We have therefore the

THEOREM 9A. *$m_i[f_i]$, $i>0$, equals $f_i[m_i]$ plus a polynomial in $f_1, f_2, \cdots, f_{i-1}[m_1, m_2, \cdots, m_{i-1}]$, containing no constant or linear terms.*

We come now to the fundamental theorem.

THEOREM 9B. *If*

$$f_i(a+b) = f_i(a)+f_i(b),$$

then

$$m_i(a+b) = \sum_k m_k(a)\, m_{i-k}(b),$$

and conversely.

Define the power series $F_a(x)$, $F_b(x)$, $F_{a+b}(x)$, $M_a(x)$, $M_b(x)$, $M_{a+b}(x)$ as we formerly defined $F_a(x)$, $M_a(x)$. If now $f_i(a+b) = f_i(a)+f_i(b)$, then

$$F_{a+b}(x) = F_a(x)+F_b(x).$$

Hence

$$M_{a+b}(x) = e^{F_{a+b}(x)} = e^{F_a(x)+F_b(x)} = e^{F_a(x)}\, e^{F_b(x)} = M_a(x)\, M_b(x).$$

Equating coefficients gives the first half of the theorem. The second half is proved similarly.

10. **Uniqueness of T and T^{-1}.** We shall show that T and T^{-1} are the only transformations obeying Theorems 9A and 9B. Consider first T.

THEOREM 10A. *Let U be a transformation from numbers $f_1, f_2, \cdots$, to numbers $m_0, m_1, m_2, \cdots$, with the following properties.*

(1) $m_0 \equiv 1$.

(2) *Each m_i, $i>0$, equals f_i plus a polynomial in $f_1, f_2, \cdots, f_{i-1}$, containing no constant or linear terms.*

(3) *If $f_i(a)$ and $f_i(b)$ are arbitrary sets of numbers, and $f_i(a+b)$ is defined by the equation* (9.1), *then the corresponding sets $m_i(a)$, $m_i(b)$, $m_i(a+b)$ are related by the equation* (9.2).

Then U is exactly T.

If we take $i=1$, we find m_1 must equal f_1. Assuming the expressions for $m_1, m_2, \cdots, m_{i-1}$ are uniquely determined by the above conditions, we shall show that that for m_i is also.

By hypothesis, we can write

$$m_i = f_i + \sum A(\alpha_1, \alpha_2, \cdots, \alpha_{i-1})\, f_1^{\alpha_1} f_2^{\alpha_2} \cdots f_{i-1}^{\alpha_{i-1}}, \tag{10.1}$$

the sum containing no constant or linear terms; that is, in each term of Σ,

$$\alpha_1 + \alpha_2 + \cdots + \alpha_{i-1} \geqq 2. \tag{10.2}$$

Let now $f_i(a)$, $f_i(b)$ be arbitrary sets of numbers, and define $f_i(a+b)$ as in condition (3). If we put

$$M = m_i(a+b) - m_i(a) - m_i(b), \tag{10.3}$$

then by conditions (1) and (3),

$$M = m_1(a)\, m_{i-1}(b) + m_2(a)\, m_{i-2}(b) + \cdots + m_{i-1}(a)\, m_1(b), \tag{10.4}$$

which contains neither $m_i(a)$ nor $m_i(b)$. But also

$$\begin{aligned}
M &= f_i(a+b) + \textstyle\sum_3 A(\gamma_1, \cdots, \gamma_{i-1}) f_1^{\gamma_1}(a+b) \cdots f_{i-1}^{\gamma_{i-1}}(a+b) \\
&\quad - f_i(a) - \textstyle\sum_1 A(\alpha_1, \cdots, \alpha_{i-1}) f_1^{\alpha_1}(a) \quad \cdots f_{i-1}^{\alpha_{i-1}}(a) \\
&\quad - f_i(b) - \textstyle\sum_2 A(\beta_1, \cdots, \beta_{i-1}) f_1^{\beta_1}(b) \quad \cdots f_{i-1}^{\beta_{i-1}}(b) \\
&= \textstyle\sum_3 A(\gamma_1, \cdots, \gamma_{i-1}) [f_1(a) + f_1(b)]^{\gamma_1} \cdots [f_{i-1}(a) + f_{i-1}(b)]^{\gamma_{i-1}} \\
&\quad - \textstyle\sum_1 - \sum_2 \\
&= \textstyle\sum_3 A(\gamma_1, \cdots, \gamma_{i-1}) \Big[\sum_{\alpha_1} \binom{\gamma_1}{\alpha_1} f_1^{\alpha_1}(a) f_1^{\gamma_1 - \alpha_1}(b)\Big] \cdots \\
&\quad \times \Big[\sum_{\alpha_{i-1}} \binom{\gamma_{i-1}}{\alpha_{i-1}} f_{i-1}^{\alpha_{i-1}}(a) f_{i-1}^{\gamma_{i-1} - \alpha_{i-1}}(b)\Big] - \textstyle\sum_1 - \sum_2 \\
&= \sum_{\gamma_1 \cdots \gamma_{i-1}} \textstyle\sum_3 A(\gamma_1, \cdots, \gamma_{i-1}) \sum'_{\alpha_1 \cdots \alpha_{i-1}} \binom{\gamma_1}{\alpha_1} \cdots \binom{\gamma_{i-1}}{\alpha_{i-1}} \\
&\quad \times f_1^{\alpha_1}(a) \cdots f_{i-1}^{\alpha_{i-1}}(a) f_1^{\gamma_1 - \alpha_1}(b) \cdots f_{i-1}^{\gamma_{i-1} - \alpha_{i-1}}(b),
\end{aligned} \tag{10.5}$$

where in $\sum'$ all terms in which

$$\alpha_1 = \alpha_2 = \cdots = \alpha_{i-1} = 0 \quad \text{or} \quad \alpha_1 = \gamma_1,\ \alpha_2 = \gamma_2, \cdots, \alpha_{i-1} = \gamma_{i-1},$$

i. e. in which b alone or a alone appears, are left out, as these are cancelled out by terms of $\sum_1$ and $\sum_2$.

If in eq. (10.4) we express each m_k in terms of the f_l, which we can do in a unique manner, by hypothesis, the resulting expression must be identically equal to the right hand side of eq. (10.5) for all values of $f_l(a)$ and $f_l(b)$, as these values are arbitrary.

This fact is sufficient to determine $A(\gamma_1, \cdots, \gamma_{i-1})$ for all sets of γ's obeying eq. (10.2) (with α_i replaced by γ_i). For choose any such set. This is a factor of the coefficient of at least one term in (10.5). For instance, if $\gamma_p \neq 0$, it is a factor of the coefficient of

$$f_p(a) f_1^{\gamma_1}(b) \cdots f_{p-1}^{\gamma_{p-1}}(b) f_p^{\gamma_p - 1}(b) f_{p+1}^{\gamma_{p+1}}(b) \cdots f_{i-1}^{\gamma_{i-1}}(b).$$

For, if $\gamma_p > 1$, $f_p(a)$ and $f_p^{\gamma_p - 1}(b)$ are both present, and if $\gamma_p = 1$, some $\gamma_q > 0$, and $f_p(a)$ and $f_q^{\gamma_q}(b)$ are both present. The whole coefficient of this term is

$$A(\gamma_1, \cdots, \gamma_{i-1}) \binom{\gamma_1}{0} \cdots \binom{\gamma_p}{1} \cdots \binom{\gamma_{i-1}}{0} = \gamma_p A(\gamma_1, \cdots, \gamma_{i-1}).$$

As this must equal the corresponding coefficient in (10.4), $A(\gamma_1, \cdots, \gamma_{i-1})$ is determined. There is thus at most a single transformation with the given properties. But T is such a transformation; hence U is T.

To prove that T^{-1} is unique, we state a theorem corresponding to the above theorem, with the m_i and f_i interchanged. We assume of course that $m_0 \equiv 1$. Now the transformation U' given in this theorem obviously has an inverse. Moreover, this inverse is seen to obey the conditions of Theorem 10A, and it is thus T. Hence U' is T^{-1}.

11. **Extension of T to the case of two subscripts.** We now consider sets of numbers f_{ij}, $(i, j) \neq (0, 0)$, and m_{ij}, where $m_{00} = 1$. Let $R_{ij}^k(f)$ be defined just as we defined R_i^k, with the condition that the first subscripts add up to i, and the second ones add up to j. Thus for example

$$\begin{aligned} &R_{22}^1(f) = f_{22}, \quad R_{22}^2(f) = 2(f_{10} f_{12} + f_{20} f_{02} + f_{01} f_{21}) + f_{11}^2, \\ &R_{22}^3(f) = 3 f_{10}^2 f_{02} + 6 f_{10} f_{01} f_{11} + 3 f_{01}^2 f_{20}, \quad R_{22}^4(f) = 6 f_{10}^2 f_{01}^2. \end{aligned} \tag{11.1}$$

Letting $f_{ij}(a)$ be the coefficient of $x^i y^j$ in a power series $F_a(x, y)$, we define $m_{ij}(a)$ as being the coefficient of $x^i y^j$ in $e^{F_a(x,y)}$, etc. Just as before, we arrive at the transformations T and T^{-1}:

$$(11.2)\qquad m_{ij} = \sum_k \frac{1}{k!} R^k_{ij}(f) = \sum_{\substack{\sum_p p\alpha_{pq}=i \\ \sum_q q\alpha_{pq}=j}} \frac{1}{\prod_{p,q} \alpha_{pq}!} \prod_{p,q} f_{pq}^{\alpha_{pq}},$$

$$(11.3)\qquad f_{ij} = \sum_k \frac{(-1)^{k-1}}{k} R^k_{ij}(m) = \sum_{\substack{\sum_p p\alpha_{pq}=i \\ \sum_q q\alpha_{pq}=j}} \frac{(-1)^s s!}{\prod_{p,q} \alpha_{pq}!} \prod_{p,q} m_{pq}^{\alpha_{pq}},$$

where

$$s = \sum_{p,q} \alpha_{pq} - 1.$$

Thus, for example,

$$\begin{aligned} m_{22} &= f_{22} + f_{10}f_{12} + f_{20}f_{02} + f_{01}f_{21} + \frac{1}{2}f_{11}^2 \\ &\quad + \frac{1}{2}f_{10}^2 f_{02} + \frac{1}{2}f_{01}^2 f_{20} + f_{10}f_{01}f_{11} + \frac{1}{4}f_{10}^2 f_{01}^2, \end{aligned}$$

$$(11.4)\qquad \begin{aligned} f_{22} &= m_{22} - m_{10}m_{12} - m_{20}m_{02} - m_{01}m_{21} - \frac{1}{2}m_{11}^2 \\ &\quad + m_{10}^2 m_{02} + m_{01}^2 m_{20} + 2\,m_{10}m_{01}m_{11} - 1\tfrac{1}{2}\,m_{10}^2 m_{01}^2. \end{aligned}$$

As before, we prove the theorems

THEOREM 11A. *$m_{ij}[f_{ij}]$, $(i, j) \neq (0, 0)$, equals $f_{ij}[m_{ij}]$ plus a polynomial in the numbers $f_{pq}[m_{pq}]$, $p \leqq i$, $q \leqq j$, $(p, q) \neq (i, j)$, $(p, q) \neq (0, 0)$, which contains no constant or linear terms.*

THEOREM 11B. *If*

$$(11.5)\qquad f_{ij}(a+b) = f_{ij}(a) + f_{ij}(b),$$

then

$$(11.6)\qquad m_{ij}(a+b) = \sum_{p,q} m_{pq}(a)\, m_{i-p,j-q}(b),$$

and conversely.

THEOREM 11C. *The transformations T and T^{-1} in two subscripts are unique in the same sense that the transformations in one subscript were.*

The transformation in one subscript can be obtained from that in two by making the second subscript always zero. The theory of the transformation holds of course with any number of subscripts. It is the case of two subscripts we shall use.

IV. THE NUMBERS f_{ij}.

12. Definition of the f_{ij}. In this part we let the elements $a, b, c, \cdots$, of Part III be graphs $G_1, G_2, G_3, \cdots$. If G_1 and G_2 are any two graphs, let $G_1 + G_2$ be their sum.

THEOREM 12A. *Given any two graphs G_1 and G_2,*

$$m_{ij}(G_1+G_2) = \sum_{p,q} m_{pq}(G_1)\, m_{i-p,j-q}(G_2). \tag{12.1}$$

For let H_1 be any subgraph of G_1 of rank p, nullity q, and let H_2 be any subgraph of G_2 of rank $i-p$, nullity $j-q$. Then H_1+H_2 is a subgraph of G_1+G_2 of rank $p+(i-p)=i$, nullity $q+(j-q)=j$. Conversely, any subgraph of G_1+G_2 of rank i, nullity j is made up of a subgraph H_1 of G_1 and a subgraph H_2 of G_2 (either of which may be the null graph), whose ranks and nullities add up to i and j respectively. From this the theorem follows.

Definition. Given any graph G, we define the numbers $f_{ij}(G)$, $(i,j) \neq (0,0)$ by the transformation T^{-1}:

$$f_{ij} = \sum_k \frac{(-1)^{k-1}}{k} R_{ij}^k(m). \tag{12.2}$$

As $m_{00} = 1$ always, Theorem 11B, Part III gives the

THEOREM 12B. *Given any two graphs G_1 and G_2,*

$$f_{ij}(G_1+G_2) = f_{ij}(G_1)+f_{ij}(G_2). \tag{12.3}$$

THEOREM 12C. *$f_{ij}(G)$ for any graph G is given by a unique polynomial in the numbers $N_1, N_2, \cdots$, containing no constant term.*

This follows immediately from Theorem 7A, Part II, as f_{ij} is a polynomial in the m_{ij} containing no constant term.

13. **Generalized graphs.** With any graph, or *real graph*, is associated a finite set of integers, positive or zero, $N_1, N_2, \cdots, N$. Consider now any finite set of integers, positive, negative or zero, $N_1, N_2, \cdots, N_q$, each N_i corresponding to a definite type T_i of non-separable graph. This collection of numbers we call a *generalized graph* G; we shall say that G contains N_i subgraphs of type T_i, $i = 1, 2, \cdots, q$. We may of course add to or remove from a generalized graph any number of numbers which are zero without altering it. We may consider a real graph as a special case of a generalized graph.

Consider now two generalized graphs G and G', whose numbers are N_i and N_i'. We let $G+G'$, the *sum* of these graphs, be the set of numbers

$$N_1+N_1',\ N_2+N_2',\ \cdots,\ N_q+N_q', \tag{13.1}$$

where q is large enough so that all the numbers of both G and G' are included. This holds of course if G and G' are real graphs. Similarly for the difference of two generalized graphs, or any other linear combination.

Various generalized graphs may be expressible in terms of real graphs. This is, in fact, true for all generalized graphs, as the following theorems show.

THEOREM 13A. *Given any type of non-separable graph T_i, the generalized graph G containing one subgraph of this type, and none of any other type, ($N_i = 1$, $N_j = 0$, $j \neq i$), can be expressed as a linear combination of real graphs.*

Let $T_1, T_2, \cdots, T_k$ be all types other than T_i of non-separable graphs which are subgraphs of a graph of type T_i. We arrange them in order, choosing first those with a single arc, next those with two arcs, etc. Types with the same number of arcs we arrange in any order. Then no graph G' of type T_p contains any graph G'' of type T_q as a subgraph, $q > p$.

Let G_1 be a graph of type T_i. Suppose G_1 contains c_2 subgraphs of type T_k. We then subtract from G_1 c_2 graphs of type T_k, forming the generalized graph G_2. In doing this, we have not altered the number of subgraphs of type T_i. G_2 contains therefore one subgraph of type T_i, and zero subgraphs of type T_k. Suppose G_2 contains c_3 subgraphs of type T_{k-1}. Subtract then c_3 graphs of type T_{k-1}, etc. This process gives us the required graph.

Example. Let G_{32} be the graph whose arcs are ab, ac, bc, bd, cd, and say G_{32} is of type T_{32}. G_{32} contains a quadrilateral. Let G_{31} be a quadrilateral; then $G_{32} - G_{31}$ is a generalized graph containing no quadrilateral. This graph contains two triangles. Let G_{21} be a triangle; then $G_{32} - G_{31} - 2\,G_{21}$ contains no triangles. As G_{32} contains five arcs, G_{31} contains four, and G_{21} contains three, $G_{32} - G_{31} - 2\,G_{21}$ contains $5 - 4 - 2 \cdot 3 = -5$ arcs. Let G_{10} be an arc. Then

$$G = G_{32} - G_{31} - 2\,G_{21} + 5\,G_{10}$$

is a generalized graph containing one subgraph of type T_{32}, but none of any other type.

THEOREM 13B. *Any generalized graph, whose numbers are $N_1, N_2, \cdots, N_q$, may be expressed as a linear combination of real graphs.*

Let G_i be a generalized graph containing one subgraph of type T_i and none of any other type. We need merely express each G_i in terms of real graphs, and put

$$G = N_1 G_1 + N_2 G_2 + \cdots + N_q G_q. \tag{13.2}$$

14. The numbers f_{ij} for generalized graphs. Given the generalized graph $G = (N_1, N_2, \cdots, N_q)$, we define $f_{ij}(G)$ by the polynomial of Theorem 12C. $m_{ij}(G)$ may of course be defined by the polynomial of Theorem 7A. It is

to be noted that, for generalized graphs, there will be in general an infinite number of non-zero f_{ij} and m_{ij}.

THEOREM 14A. *If G is a generalized graph, made up of the real graphs $G_1, G_2, \cdots, G_l$:*

$$G = G_1 + \cdots + G_k - G_{k+1} - \cdots - G_l, \tag{14.1}$$

then

$$f_{ij}(G) = f_{ij}(G_1) + \cdots + f_{ij}(G_k) - f_{ij}(G_{k+1}) - \cdots - f_{ij}(G_l). \tag{14.2}$$

By Theorem 12B, this is true if there are no minus signs present, for then G is a real graph. Assume it is true if there are fewer than $l-k$ minus signs present; we shall prove it for the case of $l-k$ minus signs.

Put

$$G' = G + G_l = G_1 + \cdots + G_k - G_{k+1} - \cdots - G_{l-1}.$$

As G' contains but $l-k-1$ minus signs,

$$f_{ij}(G') = f_{ij}(G_1) + \cdots + f_{ij}(G_k) - f_{ij}(G_{k+1}) - \cdots - f_{ij}(G_{l-1}). \tag{a}$$

Let us add to G' s graphs congruent with G_l:

$$G^{(s)} = G' + G_l + \cdots + G_l = G' + sG_l.$$

As only $l-k-1$ minus signs are used in expressing this graph in terms of real graphs, we find for $s \geq 0$

$$f_{ij}(G' + sG_l) = f_{ij}(G') + sf_{ij}(G_l).$$

Let $N_1', N_2', \cdots$, be the numbers for G', and let $N_1'', N_2'', \cdots$, be those for G_l. Then the numbers for $G' + sG_l$ are

$$N_1' + sN_1'', \ N_2' + sN_2'', \ \cdots,$$

whether s is positive or negative. Considering $N_1', N_2', \cdots, N_1'', N_2'', \cdots$, as constants, $f_{ij}(G' + sG_l)$ is a polynomial in s alone:

$$f_{ij}(G' + sG_l) = P(s),$$

with the property that

$$P(s) = f_{ij}(G') + sf_{ij}(G_l) = A + Bs$$

if s is any positive integer or zero. Therefore

$$P(s) \equiv A + Bs.$$

Now $f_{ij}(G'+s\,G_l) = P(s)$ whether s is positive or negative. Hence, putting $s = -1$, we have

$$f_{ij}(G) = A - B = f_{ij}(G') - f_{ij}(G_l)$$

which, with eq. (a), gives the desired result.

THEOREM 14B. *Let G and G' be any two generalized graphs. Then*

$$f_{ij}(G+G') = f_{ij}(G)+f_{ij}(G').$$

Express G and G' in terms of real graphs as in Theorem 14A. $G+G'$ in terms of real graphs is the sum of these expressions. $f_{ij}(G)$ is as given in Theorem 14A, and similarly for $f_{ij}(G')$, $f_{ij}(G+G')$. Comparing these three expressions gives the above equation.

The theorem obviously holds for any linear combination of generalized graphs.

THEOREM 14C. *$f_{ij}(G)$, a polynomial in the numbers $N_1, \cdots, N_q$, is linear and homogeneous.*

Let us put

$$f_{ij}(G) = P(N_1, N_2, \cdots, N_q).$$

Let G_k be a generalized graph containing one subgraph of type T_k and none of any other types, $k = 1, 2, \cdots, q$. Put

$$a_k = f_{ij}(G_k) = P(0, \cdots, 1, \cdots, 0), \tag{14.4}$$

the 1 appearing in the k'th place. Define the graph G by the equation

$$G = N_1 G_1 + N_2 G_2 + \cdots + N_q G_q,$$

the numbers $N_1, \cdots, N_q$ being any set of integers, positive, negative or zero. Then

$$f_{ij}(G) = N_1 f_{ij}(G_1) + N_2 f_{ij}(G_2) + \cdots + N_q f_{ij}(G_q)$$

by the last theorem, and thus

$$P(N_1, N_2, \cdots, N_q) = a_1 N_1 + a_2 N_2 + \cdots + a_q N_q.$$

As the polynomials on the two sides of this equation are equal for all integral values of the variables $N_1, \cdots, N_q$, they are identically equal:

$$f_{ij}(G) \equiv a_1 N_1 + a_2 N_2 + \cdots + a_q N_q, \tag{14.5}$$

and the theorem is proved.

We can, if we wish, express the polynomial $f_{ij}(G)$ in the form

$$f_{ij}(G) = a_1 N_1 + a_2 N_2 + \cdots, \tag{14.6}$$

writing down a term corresponding to every type of non-separable graph. All but a finite number of the coefficients will be zero. We can find any coefficient by the

THEOREM 14D. *The coefficient a_k of N_k in $f_{ij}(G)$ is given by eq.* (14.4).

This is true if G in terms of real graphs contains the term G_k, as we have just seen. If it does not, we must show that $a_k = 0$. Take the expression for G and add the term $0 \cdot G_k$. The proof of Theorem 14C still holds, and the right hand side of eq. (14.5) now has the added term $a_k N_k$. This can only be true if $a_k = 0$.

15. **Relation between the m_{ij} and the f_{ij}.**

THEOREM 15A. *Consider the expression for $m_{ij}(G)$ for any graph G as a polynomial in the numbers $N_1, \cdots, N_q$. $f_{ij}(G)$ is exactly the linear terms of this polynomial.*

By definition,

$$f_{ij}(G) = m_{ij}(G) + \sum_{k \geq 2} \frac{(-1)^{k-1}}{k} R_{ij}^k [m(G)]. \tag{15.1}$$

If we express the right hand side in terms of the numbers $N_1, N_2, \cdots$, we obtain the polynomial for $f_{ij}(G)$, as this expression in unique, by Theorem 12C. Hence, by Theorem 14C, all terms vanish but the linear terms. The sum on the right hand side is a polynomial in $m_{pq}(G)$ containing no constant or linear terms, and each $m_{pq}(G)$ is a polynomial in $N_1, N_2, \cdots$, containing no constant terms. Hence the sum contains no linear terms in $N_1, N_2, \cdots$, and thus the only linear terms are those in $m_{ij}(G)$, which proves the theorem.

This fact is important for the following reason.

If we know the linear terms f_{pq} of m_{pq} for all numbers $p \leqq i$, $q \leqq j$, then we can calculate m_{ij}, using the transformation T.

We shall make use of this fact in the next part.

Example. Let us find f_{30} for graphs containing no 1- or 2-circuits by eq. (14.5). Let G be a graph containing a single arc, let G' be a triangle, and let G'' be a generalized graph containing one triangle and no other subgraphs. Then

$$G'' = G' - 3G.$$

As a single arc and a triangle are the only types of graphs that can contribute to f_{30} (cf. Part II),

$$\begin{aligned} f_{30} &= f_{30}(G)E + f_{30}(G'')N_{21} \\ &= f_{30}(G)E + [f_{30}(G') - 3f_{30}(G)]N_{21}, \end{aligned}$$

if N_{21} is the number of triangles in G, by eq. (14.5). Now

$$f_{30} = m_{30} - m_{10}\, m_{20} + \frac{1}{3} m_{10}^3,$$

and

$$m_{10}(G) = 1, \quad m_{20}(G) = 0, \quad m_{30}(G) = 0,$$
$$m_{10}(G') = 3, \quad m_{20}(G') = 3, \quad m_{30}(G') = 0.$$

Hence

$$f_{30}(G) = \frac{1}{3}, \quad f_{30}(G') = -9 + 9 = 0,$$

and thus

$$f_{30} = \frac{1}{3} E - N_{21}.$$

Thence, as $f_{10} = E$ and $f_{20} = -\frac{1}{2} E$,

$$m_{30} = f_{30} + f_{10} f_{20} + \frac{1}{6} f_{10}^3 = \binom{E}{3} - N_{21}.$$

Let us find a relation between the f_{ij} analogous to eq. (1.5), Part I. To do this, we replace each term in this equation by its expression in the numbers $N_1, N_2, \cdots$ (the right hand side is already in that form). By Theorem 5A, Part II, this equation, being true for any graph, is an identity. Therefore in particular, the linear terms on the left hand side equal the linear terms on the right hand side; that is,

$$f_{i0} + f_{i-1,1} + \cdots + f_{0i} = \frac{(-1)^{i-1}}{i} E. \tag{15.2}$$

16. Independence of the m_{ij} and f_{ij}. We give here some theorems on the m_{ij} and f_{ij} similar to Theorems 5A and 5B, Part II, on the N_i.

THEOREM 16A. *If we except the numbers m_{i0}, $i \neq 1$, there is no polynomial in the remaining m_{ij} which equals 0 for all graphs and yet is not identically 0.*

If we consider only graphs with no 1- or 2-circuits, we consider only m_{ij} which are not 0 for such graphs. (Thus for instance, we do not consider m_{11}.)

Suppose there were such a polynomial. Consider then such a polynomial P in the least number of variables. Of all the numbers m_{kl} appearing, let m_{ij} be that variable, chosen from the set for which the first subscript is as great as possible, for which the second subscript is the greatest. Let us write the polynomial in the form

$$P = Q_0\, m_{ij}^{\alpha} + Q_1\, m_{ij}^{\alpha-1} + \cdots + Q_{\alpha-1}\, m_{ij} + Q_{\alpha},$$

where the Q's are polynomials in the remaining m_{kl}, and Q_0 is not zero for all graphs.

Express now each of the m_{kl} in terms of the numbers $N_1, N_2, \cdots$. By Lemma F of § 8, Part II, there is a non-separable graph of rank i, nullity j, which contributes to m_{ij} with the coefficient 1; it contributes to none of the other m_{kl}. Let us say there are N_{ij} subgraphs of this type in a given graph G. Then if N_p is the set of numbers excepting N_{ij},

$$m_{ij} = N_{ij} + R(N_p).$$

Let us put

$$Q_s(m_{kl}) = Q'_s(N_p),$$

where Q'_s does not contain N_{ij}. Now

$$\begin{aligned} P &= Q'_0[N_{ij}+R]^\alpha + Q'_1[N_{ij}+R]^{\alpha-1} + \cdots + Q'_\alpha \\ &= Q'_0 N_{ij}^\alpha + (\alpha Q'_0 R + Q'_1) N_{ij}^{\alpha-1} + \cdots, \end{aligned}$$

and N_{ij} appears only where shown. By Theorem 5A, Part II, this polynomial, being 0 for all graphs, must vanish identically. Hence the coefficients of the powers of N_{ij}, in particular, Q'_0, must vanish identically. Hence $Q_0 = Q'_0$ equals zero for all graphs. We have here a contradiction, and the theorem is proved.

THEOREM 16B. *Theorem* 16A *holds if the words "for all graphs" are replaced by the words "for all non-separable graphs".*

We merely use Theorem 5B instead of Theorem 5A.

THEOREM 16C. *Theorems* 16A *and* 16B *hold if "m_{ij}" is replaced by "f_{ij}".*

We replace "m_{ij}" by "f_{ij}" throughout the proofs.

We of course cannot include m_{i0} or f_{i0}, $i \neq 1$, in these theorems, as they are determined from the other m_{kl} or f_{kl} by eq. (1.5) of Part I or eq. (15.2) of this part.

17. **On inequalities.** The theorems of the last article tell us nothing about the existence or non-existence of inequalities on the m_{ij} or f_{ij}. There are many such inequalities, as we now show.

A fundamental inequality is the obvious one:

$$m_{ij} \geqq 0. \tag{17.1}$$

Another set of inequalities is given by eq. (2.1), Part I. Again, the number of ways of coloring any graph is $\geqq 0$:

$$\sum_i \lambda^{V-i} \sum_j (-1)^{i+j} m_{ij} \geqq 0 \tag{17.2}$$

if λ is an integer $\geqq 0$.

Inequalities on the m_i may be deduced from the interpretation of the number $(-1)^i m_i$ in terms of the broken circuits of the graph. We give one example:

$$|m_{i-1}| \geqq \frac{i}{E-i+1} |m_i|. \tag{17.3}$$

To prove this, take each of the $|m_i|$ subgraphs of G of i arcs containing no broken circuit, and drop out any arc. We form thus subgraphs of $i-1$ arcs containing no broken circuit in $i|m_i|$ ways. Consider any such subgraph. It was formed by dropping out one of the $E-(i-1)$ arcs in the rest of G, and was thus formed at most $E-i+1$ times. From this the inequality follows.

Further, some polynomials in the f_{ij} (or m_{ij}) equal a number N_{kl}, which is of course $\geqq 0$. For instance, for graphs containing no 1- or 2-circuits,

$$f_{31} = -3N_{21} + N_{31}, \qquad f_{21} = N_{21}.$$

Therefore

$$N_{31} = f_{31} + 3f_{21} \geqq 0. \tag{17.4}$$

Of course any inequality on the f_{ij} gives rise to an inequality on the m_{ij}, and conversely.

A fundamental problem, and one of extreme difficulty, is that of discovering all the inequalities.

V. Calculation of the f_{ij} and m_{ij}.

18. **Graphs containing no 1- or 2-circuits.** In such graphs, $m_{01} = m_{11} = 0$, or, $f_{01} = f_{11} = 0$. If there is but a single type of non-separable graph of rank i, nullity j, we shall call this type T_{ij}; if there are several, we shall call them ${}^1T_{ij}$, ${}^2T_{ij}$, $\cdots$, ${}^qT_{ij}$. Define N_{ij}, or ${}^1N_{ij}$, $\cdots$, ${}^qN_{ij}$ correspondingly. If there is more than one T_{ij}, put $N_{ij} = \sum_k {}^kN_{ij}$. Thus N_{ij} is the number of non-separable subgraphs of rank i, nullity j. The non-separable graphs of ranks one to four are given in the figure.

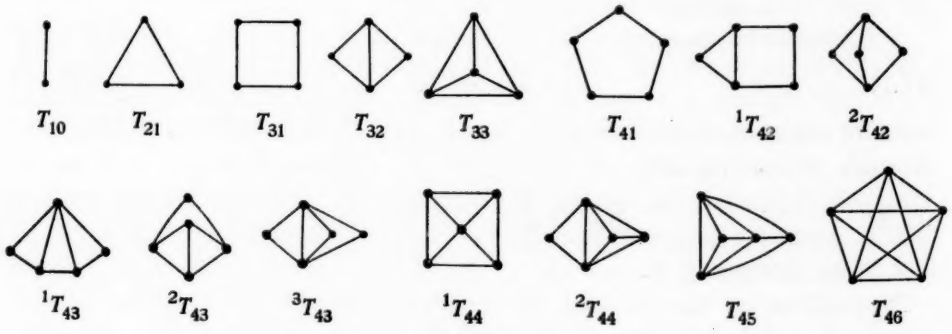

Fig. 2.

We shall calculate the f_{ij} in terms of the ${}^kN_{pq}$ in the following order: f_{10}, f_{20}, f_{21}, f_{30}, f_{31}, $\cdots$, f_{40}, $\cdots$. To find an f_{ij}, having found the

previous f_{kl}, we first write down m_{ij} in terms of the f_{kl} by eq. (11.2), Part III. We know all the f_{kl} occurring in this expression except f_{ij}. Now, by eq. (14.5), Part IV,

$$f_{ij} = a_{i_1} N_{i_1} + a_{i_2} N_{i_2} + \cdots + a_{i_p} N_{i_p},$$

where the corresponding types $T_{i_1}, \cdots, T_{i_p}$ are restricted as seen in § 8, Part II. Let $G_{i_1}, \cdots, G_{i_p}$ be non-separable graphs of these types, arranged in such an order that no one contains any one following as a subgraph. Consider first G_{i_1}, and let its numbers be $N_1(G_{i_1})$, $N_2(G_{i_1})$, $\cdots$, $N_q(G_{i_1})$. Substitute these values in the expression for m_{ij}. $m_{ij}(G_{i_1})$ will be 0 if G_{i_1} is not of rank i, nullity j, and $N_{i_2}, \cdots, N_{i_p}$ are all 0 for G_{i_1}. Hence in the resulting expression a_{i_1} is the only unknown, and we can solve for it. We can now find a_{i_2}, using G_{i_2}, etc.

Several facts will aid us. If G_{i_k} is of rank i, nullity j, then $a_{i_k} = 1$. If $j > 0$, then a_{10}, the coefficient of $E = N_{10}$, is 0 (see Theorem 8A, Part II). f_{i0}, $i > 1$, is always most simply determined by eq. (15.2), Part IV. We turn now to the calculations.

[1,0] $$m_{10} = E, \qquad f_{10} = E.$$

[2,0] $$f_{20} + f_{11} + f_{02} = f_{20} = -\frac{1}{2} E.$$

[2,1] $$m_{21} = N_{21}, \qquad f_{21} = N_{21}.$$

[3,0] $$f_{30} + f_{21} = \frac{1}{3} E, \qquad \therefore f_{30} = \frac{1}{3} E - N_{21}.$$

[3,1] $$m_{31} = f_{10} f_{21} + f_{31} = E N_{21} + (c N_{21} + N_{31}).$$

For a graph G_{21} of type T_{21},

$$m_{31} = 0, \quad E = 3, \quad N_{21} = 1, \quad N_{31} = 0.$$
$$\therefore \; 0 = 3 \cdot 1 + c \cdot 1, \quad c = -3.$$
$$\therefore \; f_{31} = -3 N_{21} + N_{31}.$$

[3,2], [3,3] $$f_{32} = N_{32}, \qquad f_{33} = N_{33}.$$

[4,0] $$f_{40} + f_{31} = -\frac{1}{4} E, \qquad \therefore f_{40} = -\frac{1}{4} E + 3 N_{21} - N_{31}.$$

$$m_{41} = \left(\frac{1}{2} f_{10}^2 f_{21} + f_{10} f_{31} + f_{20} f_{21}\right) + f_{41}$$

[4,1] $$= \left(\frac{1}{2} E^2 N_{21} - 3\frac{1}{2} E N_{21} + E N_{31}\right) + (c N_{21} + d N_{31} + e N_{32} + N_{41}).$$

$$G_{21}\colon\ E=3,\quad N_{21}=1.$$
$$\therefore\ 0=\frac{1}{2}\cdot 3^2\cdot 1-3\frac{1}{2}\cdot 3\cdot 1+c, \qquad c=6.$$
$$G_{31}\colon\ E=4,\quad N_{31}=1.$$
$$\therefore\ 0=4+d. \qquad d=-4.$$
$$G_{32}\colon\ E=5,\quad N_{21}=2,\quad N_{31}=1,\quad N_{32}=1.$$
$$\therefore\ 0=25-35+5+12-4+e, \qquad e=-3.$$
$$\therefore f_{41}=6N_{21}-4N_{31}-3N_{32}+N_{41}.$$

Similarly,

$$f_{42}=-\frac{1}{2}N_{21}-6N_{32}-6N_{33}+N_{42},$$
$$f_{43}=-6N_{33}+N_{43},\quad f_{4j}=N_{4j}, \qquad j=4,5,6;$$
$$f_{50}=\frac{1}{5}E-6N_{21}+4N_{31}+2N_{32}-N_{41},$$
$$f_{51}=-9N_{21}+10N_{31}+17N_{32}+11N_{33}-5N_{41}-3\,{}^1N_{42}-3\,{}^2N_{42}+N_{51},$$
$$f_{52}=3N_{21}+19N_{32}+36N_{33}-7\,{}^1N_{42}-6\,{}^2N_{42}-5\,{}^1N_{43}-6\,{}^2N_{43}-7\,{}^3N_{43}+N_{52}, \tag{18.1}$$
$$f_{53}=-2N_{32}+9N_{33}-9\,{}^1N_{43}-7\,{}^2N_{43}-10\,{}^3N_{43}-8\,{}^1N_{44}-8\,{}^2N_{44}+N_{53},$$
$$f_{54}=-4N_{33}-8\,{}^1N_{44}-9\,{}^2N_{44}-9N_{45}+N_{54},$$
$$f_{55}=-9N_{45}+N_{55},$$
$$f_{56}=-10N_{46}+N_{56},\quad f_{5j}=N_{5j}, \qquad j=7,8,9,10.$$

From these formulas we can write down the m_{ij}. Thence we find the m_i, the coefficients of $M(\lambda)$. The first few are

$$m_0=1,\quad m_1=-E,\quad m_2=\binom{E}{2}-N_{21},$$
$$m_3=-\binom{E}{3}+(E-2)N_{21}+N_{31}-N_{32}+N_{33}. \tag{18.2}$$

We can now find the α_i from eq. (4.4), Part I. The first six are

$$\alpha_0=1,\quad \alpha_1=0,\quad \alpha_2=-N_{21},\quad \alpha_3=-N_{31}+N_{32}-N_{33},$$
$$\alpha_4=\binom{N_{21}}{2}-3N_{33}+\sum_{j=1}^{6}(-1)^j N_{4j},$$
$$\alpha_5=N_{21}(N_{31}-N_{32}+N_{33})+6N_{33}+{}^2N_{42}-3\,{}^2N_{43}-{}^3N_{43}+4\,{}^1N_{44}+3\,{}^2N_{44}-4N_{45}+4N_{46}-\sum_{j=1}^{10}(-1)^j N_{5j}. \tag{18.3}$$

From these formulas and eq. (4.4) we can calculate the expressions for m_4 and m_5, which are rather lengthy.

19. **Regular maps.** It is well known that in studying the four color map problem, it is sufficient to restrict oneself to certain special maps; in particular, maps which we shall call regular, obeying the following conditions:[10]

(1) All the vertices are triple.

(2) There are no rings of one, two, three or four regions; that is, no set of four or fewer regions, together with all boundary lines separating them, form a multiply connected region.

(3) There are no rings of five regions except about a single region.

The numbers m_{ij} have been defined in terms of the dual graph G of the map. G', the graph consisting of the boundary lines and vertices of the map, is the dual of G in the sense defined in N. We can state the above conditions in terms of G and G' as follows:

(1′) Each vertex of G' is on exactly three arcs.

(2′) G is quintuply connected.[11] It follows that G contains no 2-circuit. For otherwise, dropping out the two corresponding arcs of G' would separate it into the pieces H_1' and H_2', each of which is seen to be of nullity >0. But from this it could be shown that dropping out the two vertices of the 2-circuit in G disconnects G.[12] Consequently G and G' determine each other uniquely.[13]

(3′) If dropping out five vertices from G disconnects G, one of the resulting pieces contains just one vertex.

The numbers $N_1, N_2, \cdots$, involved in $m_1, \cdots, m_5$, can be expressed in terms of the following quantities for all regular maps:

V = no. of regions in the map (no. of vertices in the dual graph).

$E = 3(V-2)$ = no. of boundary lines (no. of arcs).

$T = 2(V-2)$ = no. of vertices ($= N_{21}$).

Q_5 = no. of 5-sided regions (no. of vertices on five arcs).

R_6 = no. of rings of six regions, no two having a common boundary unless they are adjacent in the ring (no. of 6-circuits in which only vertices adjacent in the circuit are joined by arcs in the graph).

We derive the following formulas.

---

[10] See G. D. Birkhoff, The reducibility of maps, American Journal of Mathematics, 35 (1912), pp. 115–128.

[11] See C. G., § 2 (p. 158).

[12] Loc. cit., Theorem 9.

[13] Loc. cit., Theorems 10 and 11.

$$f_{10} = E, \quad f_{20} = -\frac{1}{2}E, \quad f_{21} = \frac{2}{3}E,$$

$$f_{30} = -\frac{1}{3}E, \quad f_{31} = -E, \quad f_{32} = E,$$

$$f_{40} = \frac{3}{4}E, \quad f_{41} = -E+Q_5, \quad f_{42} = -2\frac{1}{3}E, \quad f_{43} = 2E, \tag{19.1}$$

$$f_{50} = \frac{1}{5}E-Q_5, \quad f_{51} = 3\frac{2}{3}E+R_6, \quad f_{52} = -3E+15Q_5,$$

$$f_{53} = -6E+10Q_5, \quad f_{54} = 4\frac{2}{3}E+5Q_5, \quad f_{55} = Q_5.$$

$$m_0 = 1, \quad m_1 = -E, \quad m_2 = \binom{E}{2} - T,$$

$$m_3 = -\binom{E}{3} + (E-2)T,$$

$$m_4 = \binom{E}{4} - \binom{E-2}{2}T + \binom{T}{2} - Q_5, \tag{19.2}$$

$$m_5 = -\binom{E}{5} + \binom{E-2}{3}T - (E-4)\binom{T}{2} + (E-8)Q_5 + R_6;$$

$$\alpha_0 = 1, \quad \alpha_1 = 0, \quad \alpha_2 = -T, \quad \alpha_3 = 0,$$

$$\alpha_4 = \binom{T}{2} - Q_5, \quad \alpha_5 = 4Q_5 - R_6. \tag{19.3}$$

The regular map of the fewest number of regions is the dodecahedron, containing twelve regions, each touching five others. For this map

$$V = 12, \quad E = 30, \quad T = 20, \quad Q_5 = 12, \quad R_6 = 40.$$

We can thus calculate the first six coefficients. The whole polynomial, which was calculated by direct means, is

$$M(\lambda) = \lambda(\lambda-1)(\lambda-2)(\lambda-3)[\lambda^8 - 24\lambda^7 + 260\lambda^6 - 1670\lambda^5 + 6999\lambda^4 - 19698\lambda^3 + 36408\lambda^2 - 40240\lambda + 20170].$$

NOTE ADDED IN PROOF. R. M. Forster has calculated the m_{ij} for a large number of graphs, using the following formula. If G contains the arc ab, $G' = G - ab$, and G'' is formed from G' by letting a and b coalesce, then

$$m_{ij}(G) = m_{ij}(G') + m_{i-1,j}(G''), \tag{19.4}$$

as is easily proved.

PRINCETON UNIVERSITY.

GEODESICS ON A TWO-DIMENSIONAL RIEMANNIAN MANIFOLD WITH PERIODIC COEFFICIENTS.[1]

By Gustav A. Hedlund.

1. **Introduction.** Investigation of the properties of geodesics on closed orientable surfaces of genus one has been made in certain special cases. The geodesics on a torus have been completely classified by Bliss.[2] A study of the geodesics on a more general type of closed surface of revolution of genus one, including the torus as a special case, has been made by Kimball.[3]

The question arises as to what part the genus alone plays in determining the geodesics and whether the types of geodesics found in these special cases will be found on any surface homeomorphic with a torus. Assuming suitable conditions concerning the class of the functions defining the homeomorphism with the torus, the problem of studying the geodesics on the given surface becomes that of determining the geodesics on a two-dimensional Riemannian manifold with periodic coefficients, and can be conveniently considered in a plane. The plane is divided into congruent rectangles such that the correspondence with the original surface is one-to-one in the small, but infinity-to-one in the large. The geometry is then Euclidean in the sense that the region under consideration is a plane, not closed at infinity. This contrasts with the case of a closed orientable surface of genus greater than one where one is lead to a hyperbolic geometry in the interior of a circle.[4] In the special cases of genus one mentioned it is found that given a straight line in the plane, there is an unending geodesic of the type of this straight line, that is, a geodesic which does not wander a distance exceeding a fixed constant from the straight line. Moreover these geodesics can be so chosen that the corresponding constants are uniformly bounded. It will be shown in this paper that the same results hold in the general case. The existence of the geodesics of periodic type is known, so that what is essentially new here

[1] Received January 26, 1932.—Part of a doctoral thesis submitted at Harvard University in June, 1930.

[2] Bliss, The geodesic lines on an anchor ring, Annals of Mathematics, Second Series, 1902–1903, pp. 1–20.

[3] Kimball, Geodesics on a toroid, American Journal of Mathematics, vol. 52, 1930, pp. 29–52.

[4] Morse, A fundamental class of geodesics on any closed surface of genus greater than one, Transactions of the American Mathematical Society, vol. 26, 1924, pp. 25–60.

is the existence of those of non-periodic type. In proving the existence of these, an important property of those geodesic segments which are of Class A is derived, namely, that in the plane any such geodesic segment cannot wander a distance exceeding a fixed constant from the straight line segment joining its end points. This property makes possible the proof of a number of theorems concerning the geodesics.

In proving these theorems no use, except for positive definiteness, is made of the fact that the Riemannian manifold considered corresponds to a given surface, hence the properties derived are the properties of the geodesics on any two-dimensional Riemannian manifold with periodic coefficients provided the differential form defining the metric is positive definite.

In the attempt to extend these results to the case of more than two dimensions, the author was led to an interesting negative result showing that this case is essentially different from the case of two dimensions as far as Class A geodesics are concerned. This result, which will be proved in § 9, can perhaps be most simply stated in terms of the curves on a surface. If on a two-dimensional surface of the type considered here, a closed curve has the property that its surface length is as small as that of any closed curve into which it can be continuously deformed, the same is true of this curve taken any number of times, with respect to the closed curves into which it can be continuously deformed. The corresponding result in the case of more than two dimensions does not necessarily hold.

2. **The differential form and the geodesics.** A torus is defined as the set of points given by

$$x = \left(a + b \cdot \cos\frac{2\pi u}{\omega_1}\right)\cos\frac{2\pi v}{\omega_2}, \qquad y = \left(a + b \cdot \cos\frac{2\pi u}{\omega_1}\right)\sin\frac{2\pi v}{\omega_2},$$

$$z = b \cdot \sin\frac{2\pi u}{\omega_1}, \tag{2.1}$$

$$-\infty < u < \infty, \quad -\infty < v < \infty, \quad 0 < b < a.$$

A surface homeomorphic with a torus will then be given by

$$x = x(u, v), \qquad y = y(u, v), \qquad z = z(u, v), \tag{2.2}$$

where these functions are continuous and have the period ω_1 in u and the period ω_2 in v. It will be assumed that the functions (2.2) are of class C^4 and that at any point of the (u, v) plane at least one of the Jacobians of two of these functions with respect to u and v does not vanish.

The differential of arc length on the surface is given by

$$ds^2 = dx^2 + dy^2 + dz^2 \tag{2.3}$$

and from (2.2) this becomes

$$ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2 \tag{2.4}$$

where E, F, and G are functions of u and v of class C^3 in the (u, v) plane and have the period ω_1 in u and the period ω_2 in v. Due to the condition on the Jacobians, the form (2.4) in du and dv is non-singular and positive definite for (u, v) any point.

It will no longer be assumed that the differential form (2.4) is derived from a set of equations such as (2.2), but simply that (2.4) is given and satisfies the conditions on (2.4) concerning the class and periodicity of the coefficients and the non-singularity and positive definiteness of the form. The *geodesics* will be the curves in the (u, v) plane which are the solutions of the Euler equations corresponding to

$$J = \int \sqrt{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2}\,dt. \tag{2.5}$$

A geodesic or a geodesic segment is a curve of class C', that is, it is given parametrically by

$$u = u(t), \qquad v = v(t), \qquad \dot{u}^2 + \dot{v}^2 \neq 0 \tag{2.6}$$

where these functions are of class C^1 on some t-interval which may be infinite. A curve made up of an ordered set of oriented geodesic segments such that the final point of any one, except possibly the last, is the initial point of the following segment will be called a *broken geodesic* A broken geodesic is a geodesic if the two tangents at each of the points of juncture have the same direction.

Given an ordinary[5] curve in the (u, v) plane, its *Riemannian length*, or simply, *R-length*, will mean the value of the integral (2.5) evaluated over the given curve. Given two points, P and Q, in the (u, v) plane, it can be proved[6] that there is a curve of class C' joining P and Q, which has as small an R-length as that of any other ordinary curve joining P and Q. Moreover this curve is a geodesic segment. The R-length of such a geodesic segment will be called the *R-distance* between P and Q. The term *distance* between two points of the (u, v) plane will mean Euclidean distance.

A geodesic segment is of *class* A if it affords a minimum of R-length with respect to all ordinary curves having the same end points. From the preceding, there is at least one class A geodesic segment joining any two points of the (u, v) plane. A class A geodesic segment is a simple

[5] Bolza, Vorlesungen über Variationsrechnung, p. 192.

[6] Birkhoff, Dynamical systems with two degrees of freedom, Transactions of the American Mathematical Society, vol. 18, 1917, p. 219.

curve, for if it had a multiple point, an ordinary curve with the same end points and of less R-length could be obtained by elimination of a loop. An unending geodesic is of class A if every finite segment of it is of class A. Two class A unending geodesics cannot intersect in more than one point. For a simple proof of this statement, the reader is referred to Morse.[7]

Two unending curves will be said to be of the same *type* if there exists a constant such that, P being any point of either one of these curves, there is a point Q of the other curve such that the distance from P to Q does not exceed this constant.

3. Relation of the Riemannian distance to the Euclidean distance. Since it will be convenient to state certain theorems in terms of Euclidean distance, it is desirable to established a definite relationship between the R-distance and the distance between two points, thus making possible the statement of all such theorems in terms of R-distance.

LEMMA 3.1. *The constant A can be so chosen that*

$$E\dot{u}^2+2F\dot{u}\dot{v}+G\dot{v}^2<A^2(\dot{u}^2+\dot{v}^2) \tag{3.1}$$

for (u, v) any point of the plane, $\dot{u}$ and $\dot{v}$ arbitrary except that both shall not vanish.

Since the quadratic form on the left is positive definite for all (u, v), we have

$$EG-F^2>0, \quad E>0, \quad G>0 \tag{3.2}$$

for all values of u and v. The relationship (3.1) will hold if

$$(E-A^2)(G-A^2)-F^2>0, \quad E-A^2<0, \quad G-A^2<0, \tag{3.3}$$

for all (u, v). Let M_E be the maximum of E and M_G the maximum of G. Let A be so chosen that

$$A^2>M_E+M_G. \tag{3.4}$$

It follows with the aid of (3.2) that (3.3) holds for all (u, v) and Lemma 3.1 is proved.

By a similar method the following lemma can be proved.

LEMMA 3.2. *The constant B can be so determined that*

$$E\dot{u}^2+2F\dot{u}\dot{v}+G\dot{v}^2>B^2(\dot{u}^2+\dot{v}^2) \tag{3.5}$$

for (u, v) arbitrary, $\dot{u}$ and $\dot{v}$ subject only to the condition that they must not both vanish.

[7] Morse, loc. cit., p. 46.

THEOREM I. *The Riemannian distance between two points lies between two constant multiples of the distance, and conversely, the distance lies between two constant multiples of the Riemannian distance, the constants being independent of what pair of points is chosen.*

Let D_R and D denote the R-distance and distance, respectively, between two points P and Q. Then D_R is given by the integral (2.5) evaluated over a curve, a, of class C' joining P and Q. The distance D is not greater than the integral

$$J_E = \int \sqrt{\dot{u}^2 + \dot{v}^2}\, dt \tag{3.6}$$

evaluated over the same curve. It follows from the inequality (3.5) that

$$BD < D_R. \tag{3.7}$$

The distance between P and Q is given by the integral (3.6) evaluated over a straight line joining P and Q. The R-distance between these same points is not greater than the integral (2.5) evaluated over the same line segment. Hence from (3.1)

$$D_R < AD. \tag{3.8}$$

The inequalities (3.7) and (3.8) combined give

$$\begin{cases} BD < D_R < AD \\ \dfrac{1}{A} D_R < D < \dfrac{1}{B} D_R \end{cases} \tag{3.9}$$

and the statement of the theorem is proved.

4. **The fundamental group.** The periodicity of the coefficients E, F, and G leads to consideration of the discontinuous group of transformations

$$\begin{cases} u' = u + m\,\omega_1, \\ v' = v + n\,\omega_2, \end{cases} \qquad m, n = 0, \pm 1, \pm 2, \cdots. \tag{4.1}$$

The transformation of the group determined by the integers m and n will be denoted by $T(m, n)$. These transformations, considering the u' and v' axes as identical with the u and v axes, respectively, translate a geodesic into a geodesic. The set of points

$$0 \leqq u < \omega_1, \qquad 0 \leqq v < \omega_2,$$

will be called the *fundamental region*. The region into which the fundamental region is transformed by $T(m, n)$ will be denoted by $R(m, n)$, the fundamental region itself then being denoted by $R(0, 0)$. Two regions or two curves or two points which correspond under $T(m, n)$ will be said to be *congruent* (m, n).

A curve is *periodic with period* (m, n) if it is transformed into itself by $T(m, n)$. The period (m, n), $m \neq 0$, is a *primitive period* if m and n are relatively prime and $m > 0$. The period $(0, n)$ is a *primitive period* if $n = 1$. Given a period (m, n), there exists a primitive period (μ, ν) such-that $m = k\mu$, $n = k\nu$, where $|k|$ is the greatest common integral factor of m and n. The period (μ, ν) is the *primitive period corresponding to* (m, n).

5. **Periodic geodesics.** It will now be proved that given an arbitrary period (m, n), m and n not both zero, there is a class A periodic geodesic having this period. Any point P can be joined to the point P', congruent (m, n) to P, by a class A geodesic segment. If there are two or more such geodesic segments they will necessarily have the same R-length. The R-length of such a geodesic segment is then a single-valued function of P. This function is continuous, everywhere positive, has the period ω_1 in u, ω_2 in v, hence has a minimum greater than zero, and takes on its minimum at some point Q. Consider a class A geodesic segment b, of R-length r, joining Q and the point Q', congruent (m, n) to Q. This geodesic segment together with the geodesic segments congruent (km, kn), k any integer, form a periodic curve, g, which will be a geodesic provided there are no corners present. Corners can occur only at Q or at points congruent (km, kn), and if there is a corner, there is one at Q. But then a point T on g near Q can be joined to the point T', congruent (m, n) to T, by a segment of g of R-length r and possessing a corner. A geodesic segment joining T and T' exists which is of R-length less than r, thus contradicting the hypothesis concerning r. Thus the existence of the periodic geodesic is proved. It remains to prove the less obvious fact that the geodesic thus determined is of class A. In arriving at this result it will be necessary to prove several preliminary theorems.

A geodesic segment determined in the manner in which b was determined will be called a *basic segment.* If points which are congruent (m, n) are used in its determination, it will be called a *basic* (m, n) *segment.* The geodesic determined by a basic (m, n) segment and the geodesic segments congruent (km, kn), k any integer, will be called the *geodesic determined by the basic segment.*

As has already been noted, any class A geodesic segment, and hence in particular the basic segment b, is a simple curve. Let T be any point of b with end points Q and Q' such that Q' is congruent (m, n) to Q. Consider the geodesic segment consisting of the segment TQ' of b and the segment congruent (m, n) to the segment QT of b. This geodesic segment joins T and T', congruent (m, n) to T, has the same R-length as b and hence is a simple curve. Let the positive sense of b be that which makes Q the initial point. If the positive sense of g is determined

so that b has the same positive sense, it follows that if W is any point of g, and g is traced out in the positive sense beginning at W, along any branch if g has a multiple point at W, the point W', congruent (m, n) to W, is reached by a simple segment of g and this segment has the same R-length as the basic segment determining g. Moreover g consists of any such segment and the segments congruent (km, kn), k any integer.

It follows that g can have no multiple points. For if so, let W be such a point and let b_1 be the simple segment of g obtained by tracing g in the positive sense along one of the branches through W to the point W'. Let b_2 be the simple segment obtained by moving along another branch. The segment b_1 cannot have a point, other than the end points, in common with b_2, for then a geodesic segment of R-length less than that of b and joining points congruent (m, n) could be obtained. Now g consists of b_1 together with all geodesic segments congruent (km, kn), k any integer. If g is to contain b_2 as a subsegment, one of these geodesic segments congruent to b_1 must contain b_2 as a subsegment. But this is impossible for then the R-length of b_1 would necessarily be greater than the R-length of b_2.

THEOREM II. *Corresponding to any period (m, n), m and n not both zero, there exists a simple periodic geodesic having this period.*

Now suppose that an unending geodesic is determined by a basic (m, n) segment, where m and n are not relatively prime. What is the relationship of this basic segment with a basic (μ, ν) segment, where (μ, ν) is the primitive period corresponding to (m, n)? If $m = k\mu$, $n = k\nu$, the basic (m, n) segment is obviously of R-length not greater than $|k|$ times the R-length of a basic (μ, ν) segment. Can it be of less R-length? The answer is in the negative in the case of a two-dimensional manifold, but as shown by the example of § 9, the answer is in the affirmative in the case of more than two dimensions. It will now be proved, for the case in hand, that the basic (m, n) segment, $m = k\mu$, $n = k\nu$, consists of $|k|$ congruent basic (μ, ν) segments. The proof is suggested by considering the geodesic as on an orientable surface of genus one.

LEMMA 5.1. *A basic (m, n) segment, b, can be broken up into a finite number of geodesic segments and these can be grouped into $|k|$ groups such that if the members of a group are properly translated by transformations $T(i, j)$, each of the $|k|$ groups of segments forms a broken geodesic joining points congruent (μ, ν).*

There must be a point P, on g, such that the point P', congruent (μ, ν) to P, is also on g. For let e be the straight line joining a point Q of g and the point Q', congruent (m, n) to Q. Of the points of g on one side of e, consider those which have the property that the distance of any such point from e is as great as that of any other point of g on the

same side of e. If there are no points of g on this side of e, let the points under consideration be the points of g on e. All such points lie on a line, t_1, which has the period (m, n), and hence the period (μ, ν). Let the line t_2 be defined similarly for the points of g on the other side of e. It may happen that t_1 and t_2 coincide, in which case g is this line. If so, any point of g is such that the point congruent (μ, ν) is also on g, and the lemma is proved. Otherwise let U be a point of g on t_1. If U', congruent (μ, ν) to U, is on g, U and U' satisfy the desired condition. If U' is not on g, it lies on the boundary of a simply connected region, δ, bounded partly by a segment of t_1, and partly by a segment of g. Let V be a point of g on t_2. If V', congruent (μ, ν) to V, lies on g, the desired condition is satisfied. If not, assuming that U' does not lie on g, V' lies outside the region δ. Let g' be the geodesic congruent (μ, ν) to g. Both U' and V' lie on g'. The geodesic g' cannot cross either t_1 or t_2, hence it must cross g at some point P'. This point and the point P, congruent $(-\mu, -\nu)$ to P', are both on g. The segment of g between P and P' will be the first of the $|k|$ groups stated in the lemma.

Now consider the geodesic segment, g_2, of g, between P' and P'', congruent (m, n) to P. The end points of this segment are congruent $(m-\mu, n-\nu)$. If $|k|$ equals two, the desired division of g is attained. If $|k|$ is greater than two, let $\bar{g}$ be the curve determined by g_2 and the segments congruent $[i(m-\mu), i(n-\nu)]$, i any integer. By a repetition of the reasoning of the preceding paragraph, there is a point of $\bar{g}$ which is congruent (μ, ν) to a second point of $\bar{g}$. The segment of $\bar{g}$ between these two points constitutes the second of the desired $|k|$ groups of the lemma. If $|k|$ is three the lemma is proved. It is evident that by a continuation of this process the lemma can be proved for $|k|$ any positive integer.

With the aid of this lemma the following theorem can now be proved.

THEOREM III. *A periodic geodesic, g, of period (m, n), determined by a basic (m, n) segment, is transformed into itself by the primitive transformation corresponding to $T(m, n)$. If $T(\mu, \nu)$ is this primitive transformation, g is a geodesic determined by a basic (μ, ν) segment.*

From the lemma it was seen that the basic (m, n) segment, b, determining g was composed of $|k|$ sets of segments, $m = k\mu$, $n = k\nu$, which could be translated by transformations $T(i, j)$, thus leaving the R-length unchanged, such that each set constituted a broken geodesic with end points congruent (μ, ν). The R-length of each of these sets must then be at least r, the R-length of a basic (μ, ν) segment. But the R-length of a set cannot be greater than r, for if this were the case, b would be of R-length greater than $|k|r$. This cannot be, for a geodesic segment of R-length $|k|r$ with end points congruent (m, n) can be formed from a

basic (μ, ν) segment and $(k-1)$ congruent segments. One of the sets of the lemma was a segment of g between P and P', congruent (μ, ν) to P. It follows that this geodesic segment must be of R-length r, and hence a basic (μ, ν) segment. But a basic (μ, ν) segment has the same tangent directions at its end points, hence g is made up of this basic (μ, ν) segment and the segments congruent $(i\mu, i\nu)$, i any integer. It evidently satisfies the condition stated in the first part of the theorem.

THEOREM IV. *Two basic segments cannot have the same end points.*

For let r be the R-length of a basic (m, n) segment with end points P and P'. If there were two basic (m, n) segments joining P and P', a broken geodesic of R-length $2r$ could be formed joining P and P'', congruent $(2m, 2n)$ to P, with a corner at P'. Then a basic $(2m, 2n)$ segment would not necessarily be of twice the R-length of a basic (m, n) segment, which contradicts the preceding theorem.

THEOREM V. *Given a periodic straight line, there exists a class* A *unending periodic geodesic of the type of the given straight line.*

If the straight line is of period (m, n), let (μ, ν) be the primitive period corresponding to (m, n), and let g be a periodic geodesic determined by a basic (μ, ν) segment. If g is not of class A, there are on it two points, P and Q, which can be joined by a geodesic segment of R-length less than that of the segment of g joining these points. If k is a sufficiently large integer, there are two points $\overline{P}$ and $\overline{Q}$ of g, which are congruent $(k\mu, k\nu)$ and such that P and Q lie on the segment of g joining $\overline{P}$ and $\overline{Q}$. The R-length of the segment $\overline{P}\,\overline{Q}$ of g is kr, if r is the R-length of a basic (μ, ν) segment. But if the segment PQ of g can be replaced by a curve of less R-length, it follows that $\overline{P}$ and $\overline{Q}$ can be joined by a geodesic segment of R-length less than kr. This contradicts Theorem III. This proves Theorem V.

Now let g be a class A periodic geodesic with period (m, n). Then g must be the geodesic determined by a basic (m, n) segment. For let P be any point of g and let P' be the point congruent (m, n) to P, and hence on g. Suppose that the segment of g between P and P' is not a basic (m, n) segment and is of R-length s. The R-length, r, of a basic (m, n) segment, QQ', is then less than s. The points P and Q can be joined by a geodesic segment of R-length t. Let k be an integer, so chosen that $k(s-r) > 2t$. The points Q and Q'', congruent (km, kn), can be joined by a geodesic segment of R-length kr. Let P'' be the point into which P is transformed by the transformation which carries Q into Q''. Since P and Q can be joined by a geodesic segment of R-length t, the same is true of the points P'' and Q''. Hence the points P and P'' of g can be joined by a broken geodesic passing through Q and Q'' such that

the R-length of this broken geodesic is $kr+2t<ks$. This contradicts the hypothesis that g is of class A.

THEOREM VI. *A class* A *periodic geodesic with period* (m, n) *consists of a basic* (m, n) *segment and the segments congruent* (im, in), i *any integer.*

6. **A uniform property of class A periodic geodesics.** From Theorem VI, any class A periodic geodesic of period (m, n) is composed of basic (m, n) segments. The straight lines, t_1 and t_2, described in the preceding paragraph and associated with a class A periodic geodesic g will be called the bounding tangents of g. The only points of g on these straight lines are points of tangency, and all points of g are either on these lines or in the region between them.

Let e_u and e_v be class A periodic geodesics of the types of the u- and v-axes respectively. The first is made up of basic $(1, 0)$ segments and the second of basic $(0, 1)$ segments. Denoting by $e_u(m, n)$ the geodesic congruent (m, n) to e_u, the geodesics $e_u(0, k)$, k any integer, cannot intersect. For if two members of this family did intersect, from the periodicity they would intersect in an infinite number of points, which is impossible. Likewise no two members of the family $e_v(k, 0)$ can intersect.

These two families divide the (u, v) plane into a network of congruent regions. The boundary of any one of these regions consists of four basic segments, two of which are basic $(1, 0)$ segments, the other two being basic $(0, 1)$ segments. A region has four vertices, the extremities of the bounding basic segments. That one of the regions, the boundary of which consists of segments of $e_v(p, 0)$, $e_v(p+1, 0)$, $e_u(0, q)$, and $e_u(0, q+1)$ will be denoted by $\bar{R}(p, q)$. More precisely, $\bar{R}(p, q)$ will consist only of the interior points and those boundary points which do not lie either on $e_v(p+1, 0)$ or $e_u(0, q+1)$.

Let d be the least upper bound of the distance between any two points of one of these regions.

THEOREM VII. *The distance between the bounding tangents of any class* A *periodic geodesic* g *cannot exceed* $2d$.

Case 1. The geodesic g is of the type of the v-axis. The distance between the bounding tangents of e_v is not greater than the distance between the two points M and N on a basic segment, the tangents at which are the bounding tangents. The distance between M and N cannot exceed d, hence the same is true of the distance between the bounding tangents of e_v. The periodic geodesic $e_v(k, 0)$ is congruent to e_v, and its bounding tangents have the same property. The transformation $T(1, 0)$ moves a line which is parallel to the v-axis to the right a distance ω_1, which is not greater than d. The set of left bounding tangents of the geodesics $e_v(k, 0)$ thus divide the plane into strips of width not greater than d.

From the hypothesis that g is periodic and of the type of the v-axis, the bounding tangents of g are parallel to the v-axis. Let us suppose that the distance between the bounding tangents of g is $2d+\varepsilon$, $\varepsilon>0$. At least one of the left bounding tangents of the geodesics $e_v(k, 0)$ lies to the right of the left bounding tangent of g, and at a distance not exceeding $d+\varepsilon/2$. The right bounding tangent of the geodesic $e_v(i, 0)$ thus determined lies to the left of the right bounding tangent of g. The geodesic $e_v(i, 0)$ lies entirely between the bounding tangents of g, and must intersect g in an infinite number of points, which is impossible.

Case 2. The geodesic g is of the type of a line of positive slope. The *positive side of a straight line, not parallel to the v-axis,* will be defined as that side which contains points on the v-axis for which v is arbitrarily large and positive. The *positive side of a line parallel to the v-axis* will be that side which contains points for which u is arbitrarily large and positive.

A class A periodic geodesic g has two distinct bounding tangents except in the special case when g is a straight line. If g is not a straight line, and not of the type of the v-axis, the *positive bounding tangent* of g will be that one of the bounding tangents which has the greater v-intercept. If g is not a straight line, and is of the type of the v-axis, the *positive bounding tangent* will be that one of the bounding tangents which has the greater u-intercept. The other will be the *negative bounding tangent.* If g is a straight line, g itself is both the *positive and negative bounding tangent.*

The *positive side* of g will be that side, not including g, which has points in common with the positive side of the positive bounding tangent of g. The *negative side* will be the other.

LEMMA 6.1. *The interior points of the region $\bar{R}(p, q)$ lie on the positive side of $e_v(p, 0)$ and on the positive side of $e_u(0, q)$.*

By definition, the interior points of $\bar{R}(p, q)$ form a region bounded by the segments $e_v(p, 0)$, $e_v(p+1, 0)$, $e_u(0, q)$ and $e_u(0, q+1)$. The curve $e_v(p+1, 0)$ cannot intersect $e_v(p, 0)$, hence $e_v(p+1, 0)$ lies either on the positive or negative side of $e_v(p, 0)$. The positive bounding tangent, $b^+_{p+1,0}$, of $e_v(p+1, 0)$ is obtained from the positive bounding tangent, $b^+_{p,0}$, of $e_v(p, 0)$, by a translation parallel to the u-axis and to the right. It follows that $b^+_{p+1,0}$ lies on the positive side of $b^+_{p,0}$, and thus on the positive side of $e_v(p, 0)$. But $e_v(p+1, 0)$ has at least one point on $b^+_{p+1,0}$, and hence all points of $e_v(p+1, 0)$ lie on the positive side of $e_v(p, 0)$. Likewise $e_u(0, q+1)$ lies on the positive side of $e_u(0, q)$.

Any interior point of $\bar{R}(p, q)$ can be joined to some point of $e_v(p+1, 0)$ without crossing $e_v(p, 0)$. Since $e_v(p+1, 0)$ lies on the positive side of $e_v(p, 0)$, it follows that any interior point of $\bar{R}(p, q)$ lies on the positive

side of $e_v(p, 0)$. Similarly, any interior point lies on the positive side of $e_u(0, q)$, and the lemma is proved.

Now suppose that the distance between the tangents bounding g is $2d+\varepsilon$, $\varepsilon>0$. Let M be a point of g on the positive bounding tangent, and N a point of g on the negative bounding tangent. The point M lies in some region $\bar{R}(p, q)$ and the point N in some region $\bar{R}(s, t)$. Since the distance between the bounding tangents of g is greater than $2d$, the distance between M and N is greater than $2d$, and they cannot lie in regions which have a common interior or boundary point. The point M lies on the positive side of the negative bounding tangent of g, and is at a greater distance than $2d$ from this line. Since M is in the region $\bar{R}(p, q)$, no point of the region $\bar{R}(p+1, q-1)$ is at a distance greater than $2d$ from M, and all points of $\bar{R}(p+1, q-1)$ must lie on the positive side of the negative bounding tangent of g. From the lemma, M lies on, or on the positive side of, $e_u(0, q)$, and on the negative side of $e_v(p+1, 0)$.

Since the slope of the bounding tangents of g is positive, there is some point, P_1, of g which lies on the negative side of both $e_u(0, q)$ and $e_v(p+1, 0)$. There is a second point P_2 of g which lies on the positive side of both $e_u(0, q)$ and $e_v(p+1, 0)$. Since P_1 is on the negative side of $e_u(0, q)$ and M is on, or on the positive side of, $e_u(0, q)$, g must cross $e_u(0, q)$. Since both P_1 and M are on the negative side of $e_v(p+1, 0)$, the point at which g crosses $e_u(0, q)$ must likewise be on the negative side of $e_v(p+1, 0)$. Moreover, this is the only point in which g crosses $e_u(0, q)$. Similarly, g crosses $e_v(p+1, 0)$ just once and on the positive side of $e_u(0, q)$. Thus g can have no points in or on the boundary of the *quadrant* which consists of the points simultaneously on the positive side of $e_v(p+1, 0)$ and on the negative side of $e_u(0, q)$. For if so, g would cross either $e_v(p+1, 0)$ or $e_u(0, q)$ twice, which is impossible. Since the slope of the bounding tangents of g is positive, the points of this quadrant and its boundary lie on the negative side of g. The region $\bar{R}(p+1, q-1)$ consists of interior or boundary points of this quadrant and likewise lies on the negative side of g.

The transformation $T(p+1-s, q-1-t)$ takes the point N into a point N' in the region $\bar{R}(p+1, q-1)$. The point N' lies on the negative side of g and on the positive side of the negative bounding tangent of g. Denoting by g' the geodesic into which g is thereby transformed, the negative bounding tangent of g' has one point, N', therefore all points, on the positive side of the negative bounding tangent of g. It follows that the positive bounding tangent of g and hence M', the image of M, lies on the positive side of the positive bounding tangent of g. The point M' lies on the positive side of g and the segment $M'N'$ of g' must

cross g at some point Q. From the periodicity g' must cross g in an infinite number of points, which is impossible.

Cases 3 and 4. The geodesic g is of the type of a straight line of negative or zero slope. The proofs in these cases are so similar to the proofs already given that they may well be omitted.

7. A fundamental class of class A unending geodesics.

LEMMA 7.1. *There exists a constant D such that no class A geodesic segment can wander a distance greater than D from the straight line segment joining its end points.*

The end points, P_1 and P_2, of a class A geodesic segment, t, lie in regions $\bar{R}(a, b)$ and $\bar{R}(c, d)$, respectively. It can be assumed without loss of generality that if $a \neq c$, then $a < c$.

Case 1. $a \neq c$. $0 < (d-b)/(c-a) \leqq 1$. There is a class A unending periodic geodesic g of period $(c-a, d-b)$ which passes through $\bar{R}(a-1, b+1)$ and $\bar{R}(c-1, d+1)$. As was proved in paragraph 6, $\bar{R}(a, b)$ and $\bar{R}(c, d)$ lie on the negative side of g, and t cannot cross g, for if so, it would cross it in two points. Similarly, t cannot cross g', congruent $(2, -2)$ to g, and t lies entirely on the positive side of g'.

A point M on the positive bounding tangent of g is at a distance not greater than $4d$ from the point M', congruent $(2, -2)$ to M, and on the positive bounding tangent of g'. From Theorem VII, the point M' is at a distance not greater than $2d$ from the negative bounding tangent of g'. It follows that the distance between the positive bounding tangent of g and the negative bounding tangent of g cannot exceed $6d$. From the assumption that the slope of these bounding tangents is less than one, the vertical distance between them is not greater than $6d\sqrt{2}$.

Since $c > a$, the point P_2 lies on the positive side of $e_v(a, 0)$ and the point P_1 lies either on or on the positive side of the same geodesic. No point of t can lie on the negative side of $e_v(a, 0)$, or on the negative side of the negative bounding tangent of $e_v(a, 0)$. Similarly t cannot have a point on the positive side of the positive bounding tangent of $e_v(c+1, 0)$.

If through a point Q, of t, whose u-coördinate is greater than or equal to that of P_1, and less than or equal to that of P_2, a line is drawn parallel to the v-axis, it will intersect the line P_1P_2 in a point Q', and the distance QQ' cannot exceed $6d\sqrt{2}$. This distance is at least as great as the distance of Q from P_1P_2.

The point P_1 lies on the negative side of the positive bounding tangent of $e_v(a+1, 0)$ and either on or on the positive side of the negative bounding tangent of $e_v(a, 0)$. The distance between these tangents is not greater than d, from the definition of d. A point of t whose u-coördinate is less than that of P_1 cannot be at a greater distance from P_1 than

$$D = \sqrt{d^2 + (6d\sqrt{2} + d)^2} > 6d\sqrt{2}.$$

The same is true for the distance from a point of t with u-coördinate greater than that of P_2 to the point P_2. The lemma is proved for Case 1.

The proofs in the remaining cases:

Case 2. $c \neq a$, $1 < (d-b)/(c-a)$; Case 3. $c \neq a$, $-1 \leqq (d-b)/(c-a) < 0$;
Case 4. $c \neq a$, $(d-b)/(c-a) < -1$; Case 5. $c = a$, or $d = b$, or both;

are similar and can well be omitted.

This is the fundamental lemma needed and with its aid the following theorem can now be proved.

THEOREM VIII. *Corresponding to any straight line, there exists at least one class A geodesic of the type of this line. Conversely, every unending geodesic of class A is of the type of one straight line.*

An *element* is defined as a point and a direction through the point. An *element on a curve* of class C′ in the (u, v) plane is a point of the curve and the direction of the tangent at that point. An element has four coördinates (u, v, α, β), where u and v are the coördinates of the point and α and β are the direction cosines of the given direction. The elements (u, v, α, β) and $(u, v, -\alpha, -\beta)$ will be considered identical. A set of elements (u, v, α, β) will be said to have a *limit element* $(u_0, v_0, \alpha_0, \beta_0)$ if (u, v, α, β) considered as a point in four dimensions has $(u_0, v_0, \alpha_0, \beta_0)$ as a limit point.

A geodesic g will be said to be a *limit geodesic* of a set of geodesics, g_n, not including g, if every element on g is the limit of a sequence of elements on the members of the set g_n. The proof of Theorem VIII requires the fact that if a geodesic g is the limit geodesic of a set of class A geodesics, then g is itself of class A. For a proof of this proposition the reader is referred to Morse.[8]

Given a straight line L, let

$$\cdots P_{-2}, P_{-1}, P_0, P_1, P_2 \cdots$$

be an ordered sequence of points on L such that both P_0P_{-n} and P_0P_n become infinite with increasing n. Let g_n be a class A geodesic segment joining P_{-n} and P_n. Let K be the straight line passing through P_0 and perpendicular to L. The end points of any g_n lie on opposite sides of K, and g_n must cross K in some point Q_n. From Lemma 7.1, none of the points Q_n can be at a distance greater than D from P_0. The set of elements E_n, on g at Q_n, must have a limit element E at some point Q whose distance from P_0 does not exceed D. The element E determines

[8] Morse, loc. cit., p. 44.

an unending geodesic g, any segment of which is the limiting geodesic segment of a sequence of class A geodesic segments. Hence g is of class A.

The geodesic g thus determined is of the type of the straight line L. For from Lemma 7.1, no point of any one of the g_n lies at a distance greater than D from the line L. It follows that no point of g can lie at a distance from L exceeding D. For if this were not the case, since every point of g is a limit point of points of the g_n, some point of one of the g_n would lie at a distance from L exceeding D and we arrive at a contradiction. Thus g is confined to the region ϱ consisting of all points at a distance not exceeding D from L. This region consists of the points in and on the boundary of a strip determined by two lines, L_α and L_β, parallel to L and at a distance apart $2D$. It remains to prove that if T is any point of L, there is a point of g at a distance not exceeding D from T. This will be true if the line L', perpendicular to L at T, intersects g. Since g is unending and of class A, it is easy to prove that g, traced out in either sense from any point of g, cannot remain in a bounded region of the plane. Suppose that g did not cross L' and that L' intersects L on the interval $P_{-i}P_{-i-1}$, $i<0$. Then g would be confined to a half-strip and if traced out in either sense it would intersect all the line segments cut off by L_α and L_β on the lines L_j perpendicular to L at P_j, $j>1$. Let M_j and N_j be points of g on L_j such that the segment of g between these points passes through Q. Evidently if j is sufficiently large, these points can be joined by a geodesic segment of less R-length than the segment of g between them. It follows that g is of the type of L.

Conversely, given a class A unending geodesic g there exists a straight line L such that g is of the type of L. Let

$$\cdots P_{-2}, P_{-1}, P_0, P_1, P_2 \cdots$$

be an ordered sequence of points on g such that the R-lengths of the segments $P_{-n}P_0$ and P_0P_n of g become infinite with increasing n. From Lemma 7.1, on each of the straight lines $P_{-n}P_n$ there is at least one point Q_n whose distance from P_0 does not exceed D. The set of points Q_n has at least one limit point Q, and the set of elements on the lines $P_{-n}P_n$ at Q_n has at least one limit element E at Q. Let L be the straight line with the element E at Q.

The geodesic g is of the type of the line L. For let P be any point of g. There exists an integer N such that if $n>N$, P lies on the segment $P_{-n}P_n$ of g. From Lemma 7.1, on each of the straight lines $P_{-n}P_n$, $n>N$, there will be a point T_n whose distance from P does not exceed D. As n becomes infinite, the points T_n have at least one limiting point T on the line L. The distance of P from any one of the T_n does not exceed D,

so that the same is true of the distance PT. As in the proof of the first part of the theorem, if U is any point of L, there is a point V of g such that the distance UV does not exceed D.

From Theorem I, it is possible to state Theorem VIII readily in terms of R-distance.

8. **Theorems concerning geodesics of class A.** A number of theorems concerning the geodesics on a closed orientable surface of genus greater than one can be slightly modified so as to hold here. These theorems are stated without proof, for the essential features of the proofs are given by Morse in the paper cited.

THEOREM IX. *If two straight lines intersect, corresponding geodesics of class* A *cross.*

THEOREM X. *If two geodesics of class A are of the types of parallel lines, they lie within a finite R-distance of one another.*

THEOREM XI. *No two geodesics of class* A *which are asymptotic can intersect.*

THEOREM XII. *If a set of straight lines, M, have a straight line L as a limit line, the geodesics of class* A *of the type of the lines of M will have at least one geodesic of the type of L as a limit geodesic.*

THEOREM XIII. *A geodesic g, of class* A, *which if traced out in one of its senses comes and remains at a distance less than an assignable constant from a periodic straight line L, is either a periodic geodesic of the type of L, or else is asymptotic in both its senses to periodic class* A *geodesics which are of the type of L.*

THEOREM XIV. *If two periodic geodesics of class* A *are of the same type, they are of the same length.*

This theorem is obvious from Theorems III and VI of this paper.

THEOREM XV. *A geodesic a of class* A *which is asymptotic to a periodic geodesic b of class* A *cannot cross any periodic geodesic of the type of b.*

Given two sensed geodesics, b and c, of the same type, they will be said to have *like senses* if two points on b and c respectively, can be moved in the positive senses an arbitrarily great distance on b and c without receding more than a fixed distance from one another.

THEOREM XVI. *If there exist two different periodic geodesics of class* A, *namely, b and c, of the same type, and if between them there are no other periodic geodesics of class* A, *then if like senses be assigned to b and c, there exist at least two geodesics of class* A *lying in the region between b and c, of which one geodesic is asymptotic to b in b's positive sense and to c in c's negative sense, while the other geodesic is asymptotic to b in b's negative sense and to c in c's positive sense.*

THEOREM XVII. *A geodesic of class A cannot be asymptotic in both its senses to the same periodic geodesic.*

9. **A negative theorem.** The question arises as to whether theorems analogous to those which have been proved concerning the geodesics on a two-dimensional Riemannian manifold with periodic coefficients can also be proved for the case of an n-dimensional Riemannian manifold with periodic coefficients. Let the coefficients of the differential form

$$ds^2 = \sum_{i,j=1}^{n} A_{ij}\, dx_i\, dx_j \tag{9.1}$$

be functions of class C^3 of the variables $(x_1, \cdots, x_n)$ in the entire space E of $(x_1, \cdots, x_n)$, and periodic with period ω_i in x_i, $i = 1, \cdots, n$. Also it will be assumed that the coefficients are such that the quadratic form (9.1) is non-singular and positive definite for any fixed set of values of the x_i, $i = 1, \cdots, n$.

The geodesics corresponding to (9.1) will be curves of class C' and a *broken geodesic* will be defined as in § 2. The *R-length* of an ordinary curve in E will be the value of the integral

$$J = \int \sqrt{\sum A_{ij}\, \dot{x}_i\, \dot{x}_j}\, dt \tag{9.2}$$

evaluated over the given curve. *Distance* will mean Euclidean distance. It will be assumed that if P and Q are distinct points of E there is a geodesic segment joining P and Q which has as small an R-length as that of any ordinary curve joining P and Q. The value of (9.2) taken over this geodesic segment will be called the *R-distance* between P and Q. *Class* A geodesic segments and *class* A unending geodesics will be defined as in the two-dimensional case. The condition that two curves be of the same *type* is unchanged.

The group of transformations considered here is the group of translations

$$T(N_1, \cdots, N_n):\ x_i' = x_i + N_i\, \omega_i, \qquad i = 1, 2, \cdots, n, \tag{9.3}$$

where N_i is an integer. Again the axes are considered fixed and these transformations move sets of points about. If Z stands for a set of points, the set of points into which Z is carried by $T(N_1, \cdots, N_n)$ will be denoted by $Z(N_1, \cdots, N_n)$. It will be said that Z and $Z(N_1, \cdots, N_n)$ are *congruent* $(N_1, \cdots, N_n)$. It can readily be proved that any transformation (9.3) transforms a geodesic into a geodesic. A curve is *periodic with period* $(N_1, \cdots, N_n)$ if it is transformed into itself by $T(N_1, \cdots, N_n)$.

The set of points

$$0 \leqq x_i < \omega_i, \qquad i = 1, \cdots, n, \tag{9.4}$$

will be called the *fundamental region*, and denoted by R. The symbol $R(N_1, \cdots, N_n)$ then stands for the set of points

$$N_i\,\omega_i \leqq x_i < (N_i+1)\,\omega_i, \qquad i = 1, \cdots, n. \tag{9.5}$$

If each point P is joined to the point P', congruent $(m_1, \cdots, m_n)$ to P, by a class A geodesic segment, among the segments thus determined there is at least one whose R-length is as small as that of any other. Such a segment will be called a *basic* $(m_1, \cdots, m_n)$ *segment.*

Let g be the periodic curve formed from a basic segment with end points congruent $(m_1, \cdots, m_n)$ and the geodesic segments congruent $(km_1, \cdots, km_n)$, where k is any positive or negative integer. By a method similar to that used in the case of two-dimensions, it can be readily proved that g has no corner and is thus a geodesic. It follows that there exists a periodic geodesic of the type of any given periodic straight line. In the case of two dimensions the geodesic g thus determined is of class A. That this might not be true in the case of more than two dimensions is to be expected, for the proof in the two-dimensional case depends on the fact that a simple closed curve divides the plane into two parts. We now prove that g is not of class A.

THEOREM XVIII. *The unending periodic geodesic g formed from a basic segment and congruent segments is not necessarily of class A when the number of dimensions of the manifold exceeds two.*

The example proving this theorem is given for the case of three dimensions, but the generalization to more dimensions is obvious.

Consider the differential form

$$\begin{aligned} ds^2 &= A\,du^2 + B\,dv^2 + C\,dw^2 \\ A &= (a-\cos v)^{2n} + (a-\cos w)^{2n} \\ B &= (a-\cos[w-\pi/2])^{2n} + (a-\cos[u-\pi/2])^{2n} \\ C &= (a-\cos[u-\pi])^{2n} + (a-\cos[v-\pi])^{2n} \end{aligned} \tag{9.6}$$

where n is a positive integer. If a is chosen greater than one, the form is positive definite. The coefficients are analytic for all values of (u, v, w) and have the period 2π in all variables.

Let δ be a positive number less than $\pi/8$. The constant a is then chosen, once and for all, so that

$$a - \cos\frac{\delta}{2} = 1,$$

and thus $1 < a < 2$.

Let

$$f(v) = (a-\cos v)^{2n}.$$

Since $f(\delta) > 1$, given an arbitrarily large positive constant M, an integer N can be determined such that

$$f(\delta) > M^2, \qquad n > N.$$

From the properties of $\cos v$,

(9.7) $$f(v) > M^2, \qquad n > N,$$

except in the intervals

$$|v - 2k\pi| < \delta, \qquad k = 0, \pm 1, \cdots.$$

Moreover, since

$$f(2k\pi) = (a-1)^{2n},$$

given $\varepsilon > 0$, N' can be chosen so large that

(9.8) $$f(2k\pi) < \frac{\varepsilon^2}{2}, \qquad n > N'.$$

Denoting by $\bar{N}$ an integer larger than either N or N', (9.7) and (9.8) hold simultaneously if $n \geqq \bar{N}$.

If v is replaced by w, similar conditions hold for the second term of A. It follows that given ε and M, there exists an integer $\bar{N}$ such that

(9.9) $$A > M^2, \qquad n > \bar{N},$$

except in the *tubes* defined by

$$|v - 2i\pi| < \delta, \quad |w - 2j\pi| < \delta, \qquad i, j = 0, \pm 1, \cdots,$$

and

(9.10) $$A < \varepsilon^2, \qquad n > \bar{N}$$

on the *axes* of these tubes, which are the lines

$$v = 2i\pi, \quad w = 2j\pi, \qquad i, j = 0, \pm 1, \cdots.$$

This choice of N likewise demands that

(9.11) $$B > M^2, \qquad n > \bar{N},$$

except in the tubes

$$\left|w - 2i\pi - \frac{\pi}{2}\right| < \delta, \quad \left|u - 2j\pi - \frac{\pi}{2}\right| < \delta, \qquad i, j = 0, \pm 1, \cdots,$$

and

(9.12) $$B < \varepsilon^2, \qquad n > \bar{N},$$

at any point of the axes of these tubes. Similar conditions hold for C:

(9.13) $$C > M^2, \qquad n > \bar{N},$$

except in the tubes

$$|u - (2i+1)\pi| < \delta, \quad |v - (2j+1)\pi| < \delta, \qquad i, j = 0, \pm 1, \cdots$$

and

(9.14) $$C < \varepsilon^2, \qquad n > \bar{N},$$

on the axes of these tubes.

The distance between any two points on the axes of different tubes cannot be less than $\pi/2$. The distance between any two points of different tubes cannot be less than $\pi/2 - 2\delta > \pi/4$.

Given a curve of class C', the value of the integral

$$\bar{J} = \int \sqrt{\dot{u}^2 + \dot{v}^2 + \dot{w}^2}\, dt$$

evaluated over the given curve will be called its Euclidean length and denoted by s_E. If a curve has no points in the tubes it follows from (9.9), (9.11), and (9.13) that its R-length, s_R, satisfies the condition

$$s_R \geqq M s_E. \tag{9.15}$$

If a curve is a segment of an axis of a tube, it follows from (9.10), (9.12), and (9.14) that

$$s_R \leqq \epsilon s_E. \tag{9.16}$$

For any fixed n, the coefficients of (9.6) are bounded and less than some positive number M_n^2. It then follows that for any curve of class C'

$$s_R < M_n \cdot s_E. \tag{9.17}$$

There is a basic segment b joining points congruent (1, 1, 0) and it can be assumed that its initial point Q is a point of the fundamental region. Since the Euclidean length of b is at least $2\pi\sqrt{2}$, if there are no points of b in a tube, from (9.15) the R-length of b is at least $2\pi\sqrt{2}M$. If there is a point of b in a tube, this may be considered as an end point and the other end point must lie in a congruent tube. At least the Euclidean distance $\pi/2 - 2\delta$ must have been traversed outside of the tubes, and the R-length of b is at least $(\pi/2 - 2\delta)M$. In either case the R-length of b is at least $(\pi/2 - 2\delta)M > M\pi/4$.

If the periodic geodesic g determined by b were of class A, two points P and P' on g and congruent $(q, q, 0)$, q any positive integer, could not be joined by a geodesic segment of R-length less than $qM\pi/4$. The falsity of this will now be proved.

It can be assumed that P is a point of the fundamental region. This point can be joined to some point P_1 of the line segment

$$0 \leqq u \leqq 2\pi, \qquad v = 0, \qquad w = 0,$$

by a line segment of Euclidean length not greater than $2\pi\sqrt{2}$, and hence from (9.17) of R-length not greater than $2\pi M_n\sqrt{2}$. The point P_1 can be joined to the point $P_2(2q\pi + \pi/2, 0, 0)$ by a segment of the u-axis of length not greater than $2q\pi + \pi/2$, and from (9.16) of R-length not greater

than $(2q\pi+\pi/2)\varepsilon$. The point P_2 can be joined to the point $P_3(2q\pi+\pi/2, 0, \pi/2)$ by a segment of length $\pi/2$ and hence of R-length not greater than $M_n\pi/2$. The point P_3 is on the tube axis

$$u = 2q\pi+\pi/2, \qquad w = \pi/2,$$

and can be joined by a segment of this axis to an arbitrarily chosen point P_4 of the line segment

(9.18) $$u = 2q\pi+\pi/2, \qquad w = \pi/2, \qquad 2q\pi \leqq v \leqq 2(q+1)\pi,$$

by a segment of R-length not greater than $2(q+1)\pi\varepsilon$. Finally, the point P' in $R(q, q, 0)$ can be joined to some point P_4 of (9.18) by a line segment of R-length not greater than $2M_n\pi\sqrt{2}$.

Summing these results, P can be joined to P' by the broken line $PP_1P_2P_3P_4P'$, the R-length of which is not greater than

(9.19) $$q\cdot 4\pi\varepsilon + M_n(4\pi\sqrt{2}+\pi/2)+5\pi\varepsilon/2.$$

We wish to prove that if n is chosen properly so as to restrict M and ε to certain values, q can be chosen so large that the expression (9.19) is less in value than $qM\pi/4$. First let n be so chosen that

(9.20) $$16\varepsilon < M.$$

This is evidently possible from the preceding, and then M_n is determined since it depends only on n. Then it is possible, in consideration of (9.20), to chose q so large that

$$q\cdot 16\varepsilon + (16\sqrt{2}+2)M_n + 10\varepsilon < q\cdot M.$$

It follows that P and P' can be joined by a geodesic segment of R-length less than the R-length of the segment of g joining these two points, and g is not of class A.

Without the fact that a periodic geodesic formed from basic segments is of class A, it is impossible, by the methods used here, to prove the fundamental lemma that there exists a constant D, fixed by the given differential form, such that no class A geodesic segment can wander a distance greater than D from the straight line segment joining its end points. The example would seem to indicate that such a constant does not necessarily exist. The example also indicates that there need be no more than n class A geodesics of the types of straight lines if the manifold is n-dimensional.

BRYN MAWR COLLEGE, BRYN MAWR, PA.

ON THE SINGULAR POINTS IN THE PROBLEM OF INVERSION.[1]

By William F. Osgood.

Let an arbitrary algebraic configuration of deficiency p be given. It may be thought of in the form

$$G(x, y) = 0, \tag{1}$$

where G is an irreducible polynomial. I et the functions be uniformized by means of the automorphic functions with principal circle,

$$x = \varphi(t), \qquad y = \psi(t),$$

and let $\mathfrak{F}$ be a fundamental domain of the automorphic group. Denote by $u_\alpha(t)$, $\alpha = 1, 2, \cdots, p$, the p normal integrals of the first kind, and let

$$u_\alpha(t) = \int_{\varepsilon_\alpha}^{t} \Phi_\alpha(t)\, dt, \qquad \alpha = 1, \cdots, p.$$

The functions $\Phi_\alpha(t)$ are analytic throughout the whole interior of the unit circle, $|t|<1$.

The problem of inversion may now be formulated as follows:

$$v_\alpha = u_\alpha(\eta_1) + \cdots + u_\alpha(\eta_p), \qquad \alpha = 1, \cdots, p, \tag{2}$$

where $(v) = (v_1, \cdots, v_p)$ is an arbitrary point of the (v)-space, and where it is required to determine all values of $\eta_1, \cdots, \eta_p$ which satisfy these equations. There will always be a reduced set of values η_α, each lying in $\mathfrak{F}$, if v_α is replaced by $v_\alpha + \omega_\alpha$, where $(\omega) = (\omega_1, \cdots, \omega_p)$ is a suitable period. To the points η_α correspond points (x_α, y_α) of the Riemann's surface belonging to (1), and it is the set of values $(x) = (x_1, \cdots, x_p)$ that is usually meant in speaking of the problem of inversion.

If (v) is a point of the manifold

$$\mathfrak{M}_{p-2}: \qquad v_\alpha = -u_\alpha(s_1) - \cdots - u_\alpha(s_{p-2}), \qquad \alpha = 1, \cdots, p,$$

where the $s_1, \cdots, s_{p-2}$ are chosen each arbitrarily in the unit circle, $|t|<1$, there will still be a solution of (2), but it will not be unique; some of the points η_α may be chosen arbitrarily. If (v), however, does not lie on $\mathfrak{M}_{p-2}$, the p reduced points lying in $\mathfrak{F}$ are uniquely determined. These

[1] Received February 3, 1932.—For a detailed treatment of the properties of the functions $\Phi_\alpha(t)$ cf. the Author's *Funktionentheorie* II, p. 420.

give rise to $p!$ points (η) in the (η)-space, and to the reproductions of these points under the automorphic group.

Let $\mathfrak{A}$ denote the cylindrical domain $(\mathfrak{F}, \cdots, \mathfrak{F})$ in the (η)-space, and let $\mathfrak{B}$ be its image in the (v)-space. Then each point (η) of $\mathfrak{A}$ leads to one point (v) of $\mathfrak{Y}$; but a point (v) of $\mathfrak{B}$ not lying on $\mathfrak{M}_{p-2}$ leads to $p!$ points, in general distinct, in $\mathfrak{A}$.

We propose to examine the map in the neighborhood of the singular points of the transformation. The results illustrate the value of the method of uniformization *im Großen* by means of the automorphic functions.

The Jacobian of the transformation has the value

$$J = \frac{\partial(v_1, \cdots, v_p)}{\partial(\eta_1, \cdots, \eta_p)} = \begin{vmatrix} \Phi_1(\eta_1) & \cdots & \Phi_p(\eta_1) \\ \cdot & \cdots & \cdot \\ \Phi_1(\eta_p) & \cdots & \Phi_p(\eta_p) \end{vmatrix}. \tag{3}$$

A necessary condition for a singular point consists in the vanishing of J. The locus

$$J(\eta_1, \cdots, \eta_p) = 0 \tag{4}$$

consists in part of the surfaces[2]

$$\eta_\alpha - \eta_\beta = 0, \qquad \alpha, \beta = 1, \cdots, p, \tag{5}$$

and in part of a second surface

$$\mathfrak{D}(\eta_1, \cdots, \eta_p) = 0. \tag{6}$$

Definition of associated. p points (x_α, y_α) of the algebraic curve (1) are said to be *associated* (= *verknüpft*) if an adjoint Q_{m-3} can be passed through them. A necessary condition is, that the equations

$$C_1 \Phi_1(\eta_\alpha) + \cdots + C_p \Phi_p(\eta_\alpha) = 0, \qquad \alpha = 1, \cdots, p, \tag{7}$$

admit a solution $C_1, \cdots, C_p$, in which some $C_\alpha \neq 0$, and hence the determinant (3) must vanish.

Conversely, this condition is sufficient if the η_α are all distinct. But if some of the points (x_α, y_α) coincide, the definition of associated is as follows. Let

$$\eta_1 = \eta_2 = \cdots = \eta_k.$$

Then η_1 shall satisfy the k equations

$$C_1 \Phi_1^{(\lambda)}(\eta_1) + \cdots + C_p \Phi_p^{(\lambda)}(\eta_1) = 0, \qquad \lambda = 0, 1, \cdots, k-1. \tag{8}$$

Likewise, if

$$\eta_{k+1} = \eta_{k+2} = \cdots = \eta_{k+k'},$$

---

[2] Outside the domain $\mathfrak{A}$ there will also be the surfaces

$$\eta_\beta = L(\eta_\alpha),$$

where $t' = L(t)$ is an arbitrary substitution $\neq 1$ of the automorphic group. The surface (6) shall not include any of these.

then η_{k+1} shall satisfy the k' equations

$$C_1 \Phi_1^{(\mu)}(\eta_{k+1}) + \cdots + C_p \Phi_p^{(\mu)}(\eta_{k+1}) = 0, \quad \mu = 0, 1, \cdots, k'-1; \tag{9}$$

and so on. A necessary and sufficient condition that the present system of p linear homogeneous equations in the $C_1, \cdots, C_p$ admit a solution in which some $C_\alpha \neq 0$, consists in the vanishing of the p-rowed determinant whose principal diagonal can be written in the form

$$\begin{vmatrix} \Phi_1(\eta_1) & \Phi_2'(\eta_1) & \cdots & \Phi_k^{(k-1)}(\eta_1) & \\ & \Phi_{k+1}(\eta_{k+1}) & \Phi_{k+2}'(\eta_{k+1}) & \cdots & \Phi_{k+k'}^{(k'-1)}(\eta_{k+1}) \text{ etc.} \end{vmatrix}. \tag{10}$$

More generally, we shall speak of q points on (1) as *associated* when more adjoint curves Q_{m-3} pass through them than is in general the case for that value of q. Here, α in equations (7) runs from 1 to q, and the matrix is of lower rank than is in general the case. If, in particular, η_1 counts k times, then equations (8) must be satisfied etc. Obviously $1 < q \leqq 2p-2$.

When p points $\eta_1, \cdots, \eta_p$ are associated, we shall speak of the corresponding point (η) of the (η)-space $J = 0$ as *associated.*

THEOREM 1. *Those points of* (4) *which lie on* (6), *and only such, yield points* $(\eta_1, \cdots, \eta_p) = (\eta)$ *which are associated.*

A point (η^0) which lies on (6), but on no one of the configurations (5), is a point in which the determinant (3) vanishes, no two η_α^0 being equal to each other. Such a set of values $\eta_1^0, \cdots, \eta_p^0$ is always associated. It follows, furthermore, from continuity that, if now (η^0) be a point of (6) through which planes (5) pass, this point will also be associated. For, if a point (η) is not associated, then the same is true of all points of a certain neighborhood of this point in the (η)-space.

On the other hand, a point (η^0) which lies in one or more of the planes (5), but not on (6), is not associated. Consider first the case that only one plane (5) goes through (η^0),— say

$$\eta_1 - \eta_2 = 0. \tag{11}$$

Develop J into a series proceeding according to powers of $\eta_2 - \eta_1$. The first term vanishes, and the coefficient of the second term is the determinant

$$|\Phi_1(\eta_1)\ \Phi_2'(\eta_1)\ \Phi_3(\eta_3) \cdots \Phi_p(\eta_p)|, \tag{12}$$

or

$$\begin{vmatrix} \Phi_1(\eta_1) & \cdots & \Phi_p(\eta_1) \\ \Phi_1'(\eta_1) & \cdots & \Phi_p'(\eta_1) \\ \Phi_1(\eta_3) & \cdots & \Phi_p(\eta_3) \\ \cdots & \cdots & \cdots \\ \Phi_1(\eta_p) & \cdots & \Phi_p(\eta_p) \end{vmatrix}. \tag{13}$$

This determinant does not vanish identically, and hence J admits $\eta_2 - \eta_1$ only once as a factor. Since no other surface (5) goes through (η^0), and since (η^0) does not lie on (6) it follows that J has no further factor in the point (η^0), and so, in particular, the determinant (13) does not vanish in the point (η^0). Consequently the point (η^0) is not associated.

Turning now to the general case, that (η^0) lies on several of the planes (5), but not on (6), we find that a similar treatment is possible. For simplicity, let

$$\eta_1^0 = \eta_2^0 = \eta_3^0, \qquad \eta_4^0 = \eta_5^0,$$

where $\eta_1^0, \eta_4^0, \eta_6^0, \cdots, \eta_p^0$ are all distinct. Then

$$(\eta_2 - \eta_1)\,(\eta_3 - \eta_1)\,(\eta_3 - \eta_2) \times (\eta_5 - \eta_4)$$

will be a factor of J im Kleinen at (η^0), and there will be no other factors there. Let

$$J(\eta_1, \cdots, \eta_p) = (\eta_2 - \eta_1)\,(\eta_3 - \eta_1)\,(\eta_3 - \eta_2)\,(\eta_5 - \eta_4)\,\mathfrak{G}(\eta_1, \cdots, \eta_p), \tag{14}$$

where, now, $\mathfrak{G}(\eta_1^0, \cdots, \eta_p^0) \neq 0$. Develop $\mathfrak{G}$ according to powers of $\eta_\alpha - \eta_\alpha^0$, $\alpha = 1, 2, 3$, and $\alpha = 4, 5$. The constant term is found to be the determinant

$$\left| \Phi_1(\eta_1)\, \Phi_2'(\eta_1)\, \Phi_3''(\eta_1)\, \Phi_4(\eta_4)\, \Phi_5'(\eta_4)\, \Phi_6(\eta_6) \cdots \Phi_p(\eta_p) \right|, \tag{15}$$

formed for the point (η^0). Since this determinant does not vanish, the point (η^0) is not associated. The proof is now complete.

THEOREM 2. *The function $J(\eta_1, \cdots, \eta_p)$ contains no multiple factors.*

The proof has already been given for the factors $\eta_\alpha - \eta_\beta$, and covers at once the case of the further factors $\eta_\alpha - L(\eta_\beta)$, where $t' = L(t)$ is any transformation of the automorphic group.

Let (η^0) be an arbitrary point of (6), whether on (5) or not. Consider a factor $\mathfrak{F}$ of J, irreducible in (η^0), but not pertaining to (5). Let (η') be an ordinary point of the surface

$$\mathfrak{F}(\eta_1, \cdots, \eta_p) = 0, \tag{16}$$

arbitrarily near to (η^0). Then some derivative of $\mathfrak{F}$ is different from 0 there—say, $\partial \mathfrak{F} / \partial \eta_1 \neq 0$ — and we can solve for η_1:

$$\eta_1 = f(\eta_2, \cdots, \eta_p). \tag{17}$$

Here, $\eta_2, \cdots, \eta_p$ are independent variables, and so we can find a point $(\eta_2'', \cdots, \eta_p'')$ near $(\eta_2', \cdots, \eta_p')$, in which the rank of the matrix

$$\left\| \begin{matrix} \Phi_1(\eta_2'') & \cdots & \Phi_p(\eta_2'') \\ \cdot & \cdots & \cdot \\ \Phi_1(\eta_p'') & \cdots & \Phi_p(\eta_p'') \end{matrix} \right\| \tag{18}$$

is $p-1$. Now η_1, as given by (16), is a root of the equation

$$\begin{vmatrix} \Phi_1(t) & \cdots & \Phi_p(t) \\ \Phi_1(\eta_2) & \cdots & \Phi_p(\eta_2) \\ \cdot & \cdots & \cdot \\ \Phi_1(\eta_p) & \cdots & \Phi_p(\eta_p) \end{vmatrix} = 0. \tag{19}$$

We can find a point $(\eta_2''', \cdots, \eta_p''')$ near $(\eta_2'', \cdots, \eta_p'')$ such that all roots of (19) will be distinct. Let

$$\eta_1''' = f(\eta_2''', \cdots, \eta_p'''),$$

and develop J by Taylor's Theorem about the point (η'''). The constant term vanishes, and the coefficient of the linear term in $\eta_1 - \eta_1'''$ is

$$\begin{vmatrix} \Phi_1'(\eta_1''') & \cdots & \Phi_p'(\eta_1''') \\ \Phi_1(\eta_2''') & \cdots & \Phi_p(\eta_2''') \\ \cdot & \cdots & \cdot \\ \Phi_1(\eta_p''') & \cdots & \Phi_p(\eta_p''') \end{vmatrix}. \tag{20}$$

This determinant cannot vanish, for its vanishing would mean that η_1''' is a multiple root of (19), when $\eta_\alpha = \eta_\alpha'''$, $\alpha = 2, \cdots, p$. Hence linear terms appear in the above expansion, and so the factor $\mathfrak{F}(\eta_1, \cdots, \eta_p)$ of J is simple. This completes the proof.

THEOREM 3. *The surfaces* (5) *belong to Case* I; *the surfaces* (6) *to Case* II.[3]

For, in the neighborhood of an ordinary point of (5), the inverse transformation is finitely multiple-valued; in every point of (6) it is infinitely multiple-valued.

THEOREM 4. *Let* (η^0) *be a point lying on one or more of the configurations* (5), *but not on* (6). *In the neighborhood of* (η^0) *the inverse transformation is finitely multiple-valued.*

Since the point (η^0) is not associated, the corresponding point (v^0) leads only to isolated points in the (η)-space, and hence, in particular, to no second point of the neighborhood of (η^0). It follows, then, from Theorem 2 of p. 139, *Funktionentheorie* Π_1, that η_p is given by an irreducible equation of the form:

$$\eta_p^N + A_1(v_1, \cdots, v_p)\eta_p^{N-1} + \cdots + A_p(v_1, \cdots, v_p) = 0, \tag{21}$$

where $A_\alpha(v_1, \cdots, v_p)$ is analytic in the point (v^0), and that $\eta_1, \cdots, \eta_{p-1}$ are single-valued and continuous, and in general analytic, on (21);—or conceivably the $\eta_1, \cdots, \eta_p$ must first be subjected to a linear transformation,

[3] Cf. the Author's *Funktionentheorie* II, p. 143.

the above statement then holding for the transformed $\eta_1', \cdots, \eta_p'$. Moreover, N is equal precisely to the number of points (η) that coalesce in (η^0), namely, $k_1!\, k_2! \cdots k_q!$, where

$$\eta^0_{\lambda_1} = \eta^0_{\lambda_1+1} = \cdots = \eta^0_{\lambda_1+k_1-1}, \quad \text{etc.}$$

Each point (v) of the neighborhood of (v^0) goes over into a unique point (η) of the N-leaved Riemann's surface corresponding to (21).

The full import of the generality of the foregoing theorem is shown by the following example.

An Example. Consider the transformation:

$$\begin{cases} x = v, \\ y = vw - u, \\ z = \frac{1}{2} vw^2 - uw. \end{cases} \tag{22}$$

Here,

$$J(u, v, w) = \frac{\partial(x, y, z)}{\partial(u, v, w)} = u. \tag{23}$$

It is seen that the surface $u = 0$ belongs to Case I, since that surface is mapped, in general, in a one-to-one fashion on a surface

$$x = v, \quad y = v\,w, \quad z = \tfrac{1}{2} v\,w^2, \tag{24}$$

and not on a manifold of lower dimensionality. But a curve in that surface, namely, the axis of w, goes over into a single point, the origin.

Thus in Case I, even when an irreducible factor of J, set equal to 0, yields a surface that has no singular point, it cannot be inferred that the inverse transformation is finitely multiple-valued. That, in the case of (4), this is indeed true in an ordinary point of a configuration (5) is much; that it is true of *every* point of (5) not lying on (6), is all that could be desired.

We note furthermore that, although each point of the surface $u = 0$ is carried over into a (single) point of the surface

$$\Delta(x, y, z) = y^2 - 2xz = 0, \tag{25}$$

the converse is true only in general. A point (x_0, y_0, z_0) of (25) on the line $x = 0$, $y = 0$ has no point corresponding to it unless $z_0 = 0$; and then every point of the axis of w corresponds to it.

The inverse transformation is:

$$\begin{cases} u = xw - y, \\ v = x, \\ \frac{1}{2} xw^2 - yw + z = 0. \end{cases} \tag{26}$$

The discriminant of the last equation is $\Delta(x, y, z)$.

For completeness we add the explicit statement of the following theorem, although the result is contained implicitly in the foregoing considerations.

THEOREM 5. *Each point of* (6) *is mapped on a point of* $\mathfrak{M}_{p-2}$; *each point of* $\mathfrak{M}_{p-2}$ *is mapped on an infinite number of points of* (6).

The manifolds (5), *exclusive of their points which lie on* (6), *are mapped in a one-to-one manner and continuously on a Riemann manifold, in the interior points of which at most a finite number of leaves are joined together.*

The surface (6) may be wholly free from singularities. This is the case when $p = 3$ and Noether's Normal Curve — here, the plane quartic, the hyperelliptic case being excluded — has 28 distinct double tangents. For now, every point of inflexion is a simple point, i. e. the inflexional tangent cuts the curve there in only three points. But if, as may happen, there is a point of hyperinflexion, i. e. a point in which the inflexional tangent cuts in four coincident points, the surface (6) will be singular at the corresponding point.

In general, a point of (6) which lies in no plane (5), or which lies at most in a single plane (5), will be non-singular. On the other hand, a point (η^0) of (6) not lying on any surface (5) will be a singular point of (6) if every set of $p-1$ points chosen from the $\eta_1^0, \cdots, \eta_p^0$ is associated.

Finally, let (η^0) be an arbitrary point of (6), in which

$$\eta_1^0 = \eta_2^0 = \cdots = \eta_{k_1}^0, \qquad \eta_{k_1+1}^0 = \cdots = \eta_{k_2}^0, \qquad \text{etc.},$$

where $\eta_{k_i}^0 \neq \eta_{k_i+1}^0$. Then the determinant whose first column is

$$\Phi_1(\eta_1^0),\ \Phi_1'(\eta_1^0),\ \cdots,\ \Phi_1^{(k_1-1)}(\eta_1^0),\quad \Phi_1(\eta_{k_1+1}^0),\ \Phi_1'(\eta_{k_1+1}^0),\ \cdots,\ \Phi_1^{(k_2-1)}(\eta_{k_1+1}^0)\ \cdots$$

vanishes. In order that (η^0) be a singular point, it is necessary and sufficient that each of the following determinants vanish. Replace the (k_1-1)st row by $\Phi_1^{(k_1)}(\eta_1^0),\ \Phi_2^{(k_1)}(\eta_1^0),\ \cdots,\ \Phi_p^{(k_1)}(\eta_1^0)$. Next, returning to the initial determinant, replace the row beginning with $\Phi_1^{(k_2-1)}(\eta_{k_1+1}^0)$ by a row in which each $\Phi_\alpha^{(k_2-1)}(\eta_{k_1+1}^0)$ is replaced by $\Phi_\alpha^{(k_2)}(\eta_{k_1+1}^0)$, and so continue to the end.

HARVARD UNIVERSITY,
January 1932.

ON THE WEIERSTRASS CONDITION FOR THE PROBLEM OF BOLZA IN THE CALCULUS OF VARIATIONS.[1]

By Lawrence M. Graves.

In a recent number of these Annals[2] Bliss has discussed sufficient conditions for a minimum in the problem of Bolza. For this problem, the present paper derives the necessary condition of Weierstrass by a new method, applicable to minimizing arcs which are assumed to be normal only on the whole interval $(x_1\, x_2)$ relative to the given end-conditions. For certain types of abnormal arcs, the new method yields a new necessary condition for a minimum, which is also expressed in terms of Weierstrassian E-functions, but not so simply as the well known condition of Weierstrass. The new method is suggested by the well known method due to Bliss for deriving the multiplier rule.

The problem of Bolza is that of finding in a class of arcs

$$y_i = y_i(x), \quad (i = 1, \cdots, n;\ x_1 \leqq x \leqq x_2),$$

satisfying the differential equations and end-conditions

$$\begin{aligned} \varphi_\alpha(x, y, y') &= 0, \qquad (\alpha = 1, \cdots, m < n), \\ \psi_\mu[x_1, y(x_1), x_2, y(x_2)] &= 0, \quad (\mu = 1, \cdots, p \leqq 2n+2), \end{aligned} \tag{1}$$

one which minimizes

$$I = \int_{x_1}^{x_2} f(x, y, y')\, dx + G[x_1, y(x_1), x_2, y(x_2)].$$

For historical references, notations, and a precise statement of the hypotheses, the reader is referred to the paper of Bliss already cited.

We shall set

$$\begin{aligned} I_1(\xi, \eta) = {} & \int_{x_1}^{x_2} \{f_{y_i}\eta_i + f_{y_i'}\eta_i'\}\, dx - f(x_1)\,\xi_1 + f(x_2)\,\xi_2 \\ & + \frac{\partial G}{\partial x_1}\xi_1 + \frac{\partial G}{\partial x_2}\xi_2 + \frac{\partial G}{\partial y_{i1}}\eta_i(x_1) + \frac{\partial G}{\partial y_{i2}}\eta_i(x_2), \\ \Psi_\mu(\xi, \eta) = {} & \frac{\partial \psi_\mu}{\partial x_1}\xi_1 + \frac{\partial \psi_\mu}{\partial x_2}\xi_2 + \frac{\partial \psi_\mu}{\partial y_{i1}}\eta_i(x_1) + \frac{\partial \psi_\mu}{\partial y_{i2}}\eta_i(x_2), \end{aligned}$$

and suppose that for arbitrary sets of admissible variations $(\xi_\sigma, \eta_\sigma)$ of the minimizing arc

[1] Received February 29, 1932.

[2] Bliss, The problem of Bolza in the calculus of variations, this volume pp. 261–274.

$$E_{12}\colon\ y_i = y_i(x),$$

the matrix

$$\| \Psi_\mu(\xi_\sigma, \eta\sigma) \| \qquad (\mu, \sigma = 1, \cdots, p), \tag{2}$$

has maximum rank $p-q+1$. If $q=1$, then E_{12} is normal relative to the end-conditions $\psi_\mu = 0$, and it is a necessary condition for a minimum that the matrix

$$\left\| \begin{matrix} I_1(\xi_\sigma, \eta_\sigma) \\ \Psi_\mu(\xi_\sigma, \eta_\sigma) \end{matrix} \right\| \qquad (\mu = 1, \cdots, p;\ \sigma = 0, 1, \cdots, p) \tag{3}$$

also has maximum rank p. If $q>1$, we make *the restrictive hypothesis that the matrix* (3) *has the same maximum rank as the matrix* (2). Then there exist q linearly independent sets of constant multipliers $\lambda_0^\nu, l_1^\nu, \cdots, l_p^\nu$, such that

$$\lambda_0^\nu I_1(\xi, \eta) + l_\mu^\nu \Psi_\mu(\xi, \eta) = 0, \qquad (\nu = 1, \cdots, q) \tag{4}$$

for every admissible variation (ξ, η). We may choose

$$\lambda_0^1 = 1, \quad \lambda_0^2 = \cdots = \lambda_0^q = 0. \tag{5}$$

It may be shown in the usual way[3] that with each such set there are associated multipliers $\lambda_\alpha^\nu(x)$ and constants c_i^ν such that, if

$$F^\nu = \lambda_0^\nu f + \lambda_\alpha^\nu \varphi_\alpha, \tag{6}$$

then along the minimizing arc E_{12}

$$F_{y_i'}^\nu = \int_{x_1}^{x} F_{y_i}^\nu \, dx + c_i^\nu, \tag{7}$$

$$\lambda_0^\nu G_{y_{i2}} + l_\mu^\nu \psi_{\mu y_{i2}} + F_{y_i'}^\nu(x_2) = 0. \tag{8}$$

The equations (8) are the only ones of the transversality conditions we shall need to use.

Let t_ν, $(\nu = 1, \cdots, q)$, be points interior to the interval $(x_1\, x_2)$ which do not correspond to corners of the arc E_{12}, let $y_{i\nu} = y_i(t_\nu)$, and let $Y_{i\nu}'$ be such that the sets (t_ν, y_ν, Y_ν') are all admissible. Then as usual we can adjoin additional functions[4] $\varphi_{m+1}, \cdots, \varphi_n$, with arguments x, y, y',

[3] See Bliss, "The problem of Lagrange in the calculus of variations", American Journal of Math., vol. 52 (1930), pp. 674–744; "The problem of Mayer with variable end-points", Transactions of the American Math. Soc., vol. 19 (1918), pp. 305–314; Morse and Myers, "The problems of Lagrange and Mayer with variable end-points", Proceedings of the American Academy of Arts and Sciences, vol. 66 (1931), pp. 235–253.

[4] See Bliss, The problem of Mayer with variable end-points, loc. cit.

having required continuity properties, and such that the matrix $\varphi_{iy'_j}$ $(i, j = 1, \cdots, n)$, is non-singular along E_{12} and also at the elements (t_ν, y_ν, Y'_ν). Let

$$\begin{aligned} \varphi_i(x, y, y') &= z_i(x) \quad \text{along } E_{12}, \\ \varphi_i(t_\nu, y_\nu, Y'_\nu) &= Z_{i\nu}, \\ \bar{z}_i(x, \beta, \varepsilon) &= Z_{i\nu} \quad \text{on} \quad (t_\nu, t_\nu + \beta_\nu), \\ &= z_i(x) + \varepsilon_\sigma \zeta_{i\sigma}(x) \quad \text{on the set } S. \end{aligned}$$

Here S is the complement of the set of intervals $(t_\nu, t_\nu + \beta_\nu)$, and it is understood that the first m components of the sets $z_i(x)$, $Z_{i\nu}$, $\zeta_{i\sigma}(x)$ are zero, while the $\zeta_{i\sigma}(x)$ are otherwise arbitrary functions having only a finite number of finite discontinuities. Then the differential equations and initial conditions

$$\varphi_i(x, \bar{y}, \bar{y}') = \bar{z}_i, \quad \bar{y}_i(x_1) = y_i(x_1) + \varepsilon_\sigma \eta_{i\sigma}(x_1),$$

have a unique continuous family of solutions,

$$y_i = \bar{y}_i(x, \beta, \varepsilon), \tag{9}$$

containing E_{12} for $\beta_\nu = \varepsilon_\sigma = 0$. On each of the sub-intervals into which the points t_ν, $t_\nu + \beta_\nu$ divide the interval $(x_1\, x_2)$, the functions (9) have continuous partial derivatives with respect to all their arguments. If we set

$$\Phi_i(\eta, \eta') \equiv \varphi_{iy_j}\eta_j + \varphi_{iy'_j}\eta'_j$$

we find that

$$\eta_{i\sigma}(x) \equiv \frac{\partial}{\partial \varepsilon_\sigma}\bar{y}_i(x, 0, 0), \qquad \theta_{i\nu}(x) \equiv \frac{\partial}{\partial \beta_\nu}\bar{y}_i(x, 0, 0)$$

satisfy respectively the equations

$$\Phi_i(\eta_\sigma, \eta'_\sigma) = \zeta_{i\sigma}, \tag{10}$$

$$\Phi_i(\theta_\nu, \theta'_\nu) = 0. \tag{11}$$

Since $\bar{y}_i(x, \beta, \varepsilon)$ is independent of β_ν on the interval $(x_1, t_\nu + \beta_\nu)$, we have

$$\theta_{i\nu}(x) = 0 \quad \text{on} \quad (x_1, t_\nu), \tag{12}$$

and by differentiating $\bar{y}_i(t_\nu + \beta_\nu, \beta, \varepsilon)$ with respect to β_ν we find

$$\theta_{i\nu}(t_\nu + 0) = Y'_{i\nu} - y'_i(t_\nu) = Y'_{i\nu} - y'_{i\nu}. \tag{13}$$

Now suppose that $\xi_{1\sigma}$, $\xi_{2\sigma}$, $\zeta_{i\sigma}$, $(\sigma = 1, \cdots, p - q + 1)$, are chosen so as to give the matrix (2) its maximum rank $p - q + 1$. If we substitute

$\bar{y}_i(x, \beta, \varepsilon)$, $\bar{x}_1(\varepsilon) = x_1 + \varepsilon_\sigma \xi_{1\sigma}$, $\bar{x}_2(\varepsilon) = x_2 + \varepsilon_\sigma \xi_{2\sigma}$, $(\sigma = 1, \cdots, p-q+1)$, in the expressions I and ψ_μ, we obtain $p+1$ functions $I(\beta, \varepsilon)$, $\psi_\mu(\beta, \varepsilon)$. If the functional determinant D of the equations

$$\begin{aligned} I(\beta, \varepsilon) &= I(0, 0) + u, \\ \psi_\mu(\beta, \varepsilon) &= 0 \end{aligned} \tag{14}$$

with respect to the variables $\beta_1, \cdots, \beta_q, \varepsilon_1, \cdots, \varepsilon_{p-q+1}$, does not vanish for $\beta = \varepsilon = 0$, then equations (14) have solutions $\beta_\varrho(u)$, $\varepsilon_\sigma(u)$ for u sufficiently near zero. If all the derivatives $\beta'_\varrho(0)$ are negative, then there exist admissible arcs $\bar{y}_i(x, \beta, \varepsilon)$ satisfying the end-conditions $\psi_\mu = 0$ and making the variable u in (14) negative. Hence a necessary condition for a minimum of the expression I is that if the equations

$$\begin{aligned} \frac{\partial I}{\partial \beta_\varrho}\beta'_\varrho + \frac{\partial I}{\partial \varepsilon_\sigma}\varepsilon'_\sigma &= 1, \\ \frac{\partial \psi_\mu}{\partial \beta_\varrho}\beta'_\varrho + \frac{\partial \psi_\mu}{\partial \varepsilon_\sigma}\varepsilon'_\sigma &= 0, \qquad (\mu = 1, \cdots, p), \end{aligned} \tag{15}$$

have a unique solution β'_ϱ, ε'_σ, then not all the β'_ϱ are negative. By use of equations (11), (12), (7), (1), and (13), we find that

$$\begin{aligned} \lambda_0^\nu \frac{\partial I}{\partial \beta_\varrho} = {} & F^\nu(t_\varrho, y_\varrho, Y'_\varrho) - F^\nu(t_\varrho, y_\varrho, y'_\varrho) - (Y'_{i\varrho} - y'_{i\varrho}) F^\nu_{y'_i}(t_\varrho, y_\varrho, y'_\varrho) \\ & + F^\nu_{y'_i}(x_2)\, \theta_{i\varrho}(x_2) + \lambda_0^\nu G_{y_{i_2}} \theta_{i\varrho}(x_2), \end{aligned}$$

where there is no summation with respect to the index ϱ. Also by (12) we have $\partial \psi_\mu / \partial \beta_\varrho = \psi_{\mu y_{i_2}} \theta_{i\varrho}(x_2)$. Hence if we multiply the equations (15) by the multipliers λ_0^ν, l_μ^ν, and add, we obtain by recalling equations (4), (5) and (8) the new system of equations

$$\begin{aligned} E^1_\varrho \beta'_\varrho &= 1, \\ E^\nu_\varrho \beta'_\varrho &= 0 \qquad (\nu = 2, \cdots, q) \end{aligned} \tag{16}$$

where

$$\begin{aligned} E^\nu_\varrho = {} & F^\nu(t_\varrho, y_\varrho, Y'_\varrho) - F^\nu(t_\varrho, y_\varrho, y'_\varrho) \\ & - (Y'_{i\varrho} - y'_{i\varrho}) F^\nu_{y'_i}(t_\varrho, y_\varrho, y'_\varrho), \qquad (\varrho, \nu = 1, \cdots, q). \end{aligned} \tag{17}$$

It is readily verified as a consequence of the choice of the variations $(\xi_\sigma, \eta_\sigma)$ that the determinant of the coefficients of equations (16) is a non-vanishing multiple of the determinant D.

The statement: "If equations (16) have a unique solution, not all its components β'_ϱ are negative", is equivalent to the statement: "The minors obtained from the matrix of q rows and $q+1$ columns

$$\begin{pmatrix} E_1^1 & \cdots & E_q^1 & 1 \\ E_1^2 & \cdots & E_q^2 & 0 \\ \vdots & \cdots & \vdots & \vdots \\ E_1^q & \cdots & E_q^q & 0 \end{pmatrix} \tag{18}$$

by deleting single columns cannot be all different from zero and of alternate sign."[5]

In the special case $q = 1$, the arc E_{12} is normal with respect to the end-conditions $\psi_\mu = 0$, (16) becomes the single equation $E(x, y, y', Y')\beta' = 1$, and the necessary condition reduces to the well known condition of Weierstrass, $E \geqq 0$. In case $q = 2$, the necessary condition that the two-by-two minors of the matrix (18) do not alternate in sign is equivalent to saying that either

$$E_1^2 E_2^2 \geqq 0 \tag{19}$$

or

$$E_1^1 - E_2^1 E_1^2 / E_2^2 \geqq 0 \tag{20}$$

for every t_1, Y_1', t_2, Y_2' satisfying the conditions imposed. If the inequality (19) holds true for every such t_1, Y_1', t_2, Y_2', then the arc E_{12} satisfies the multiplier rule and the necessary condition of Weierstrass for the Mayer problem of minimizing or maximizing one of the functions ψ_μ in the class of admissible arcs for which the remaining $\psi_\mu = 0$. But if inequality (19) fails for a pair of points t_1, t_2, then the class of admissible arcs satisfying the end-conditions $\psi_\mu = 0$ is certainly non-singular, and that is why an additional condition (20) may be obtained.

The preceding results are summarized in the

THEOREM. *Let E_{12} be a minimizing arc for the expression I in the class of admissible arcs satisfying the end-conditions $\psi_\mu = 0$. If E_{12} is normal relative to the end-conditions, then the Weierstrass condition*

$$\begin{aligned} E(x, y, y', \lambda, Y') = {} & F(x, y, Y', \lambda) - F(x, y, y', \lambda) \\ & - (Y_i' - y_i') F_{y_i'}(x, y, y', \lambda) \geqq 0 \end{aligned}$$

holds at every element (x, y, y', λ) of the arc E_{12}, for every Y' such that (x, y, Y') is admissible. If E_{12} is not normal, and if the matrices (2) *and* (3) *have the same maximum rank $p - q + 1$, then for every set of admissible elements $[t_1, y(t_1), Y_1'], \cdots, [t_q, y(t_q), Y_q']$, with the points $[t_\nu, y(t_\nu)]$ on the arc E_{12}, the minors obtained from the matrix* (18) *by deleting single columns cannot be all different from zero and of alternate signs. The elements E_ϱ^ν of* (18) *are defined by* (17), (6), (4) *and* (5).

[5] I am indebted for the last form of statement of the condition to L. L. Dines.

As a simple example, consider the isoperimetric problem, that is, the problem of minimizing an integral

$$I = \int f(x, y, y')\, dx$$

in the class of curves for which a second integral

$$J = \int g(x, y, y')\, dx$$

takes a prescribed constant value. Suppose the minimizing curve E_{12} is an extremal for both integrals. Denote the Weierstrass E-function for the integral I and the curve E by $E_I(x, Y')$, and for the integral J and the curve E by $E_J(x, Y')$. Then it is necessary that for every pair (t_1, Y_1'), (t_2, Y_2') for which

$$E_J(t_1, Y_1')\, E_J(t_2, Y_2') < 0$$

we have

$$E_I(t_1, Y_1') - E_I(t_2, Y_2')\, E_J(t_1, Y_1')/E_J(t_2, Y_2') \geqq 0.$$

It is plain that this condition is always verified if the Weierstrass condition for the integral I is satisfied along the curve E.

THE UNIVERSITY OF CHICAGO.

ON A THEOREM OF GROSS AND IVERSEN.[1]

By Joseph L. Doob.

One of the most powerful generalizations of the theorem of Picard and allied theorems is a theorem proved by W. Gross and F. Iversen.[2] The purpose of this paper is to show how the proof can be simplified using a theorem due to W. Seidel.[3] Before stating these theorems, certain concepts due to Painlevé will be defined.

Let $f(z)$ be a function meromorphic in a domain[4] γ, bounded by a closed Jordan curve Γ (i. e. a one to one and continuous image of the perimeter of a circle). In the following discussion, points of the extended plane, or of the Riemann sphere corresponding to it by stereographic projection, will be considered. "Closed", "open", etc., used of point sets in the plane will refer to the corresponding point sets on the sphere. The point ∞ is then in no way exceptional. Let $\{P_n\}$ be a sequence of points in γ approaching P on Γ as a limit. Then if $\lim_{n\to\infty} f(P_n) = \alpha$, α will be called a cluster value of $f(z)$ in γ at P. The set of all cluster values of $f(z)$ in γ at P is called the cluster set of $f(z)$ in γ at P. This set is closed and connected. Cluster values on given paths are defined by restricting the sequences $\{P_n\}$ to lie on those paths. If there is a continuous curve to P on which $f(z)$ has only one cluster value, this value is called a convergence value of $f(z)$ in γ at P. If $f(z)$ is continuous on Γ (except perhaps at P), the cluster set of $f(z)$ on Γ at P is defined in an obvious way. This set is the sum of two closed connected point sets—the cluster sets of $f(z)$ on Γ on each side of P. The cluster set of $f(z)$ on Γ at P can be defined even if $f(z)$ is not defined on Γ. Let $\{P_n\}$ be a sequence of distinct points on Γ in a neighborhood of P, on one side of P, such that P_{n+1} is between P_n and P and such that $\lim_{n\to\infty} P_n = P$. Let C_n be the sum of all the cluster sets of $f(z)$ in γ at the points of Γ between P_n

[1] Received January 15, 1932.

[2] W. Gross, Monatshefte für Mathematik und Physik, vol. 29 (1918), pp. 3–47; Mathematische Zeitschrift, vol. 2 (1918), pp. 242–294.

F. Iversen, Recherches sur les fonctions inverses des fonctions méromorphes, (thesis) Helsingfors 1914; Översikt av Finska Vetenskaps-Societetens Förhandlingar, vol. 58 (1915–1916), Section A, No. 25; ibid., vol. 64 (1921–1922), Section A, No. 4.

[3] W. Seidel, Transactions of the American Mathematical Society, vol. 34 (1932), pp. 3–4.

[4] In this paper, every open connected point set will be called a domain.

and P, (not including P), and let C_n' be the set C_n together with all its limit points. Then $C_n' \supset C_{n+1}'$ and the cluster set of $f(z)$ on Γ at P (on the side considered) is defined as the product of all the sets $C_1', C_2', \cdots$. This set is seen at once to be closed and connected and independent of the sequence $\{P_n\}$ chosen. The cluster set of $f(z)$ on Γ at P is then defined as the sum of the cluster sets on either side. The theorem of Seidel can now be stated as follows.

THEOREM 1. *Let $f(z)$ be analytic and bounded in a domain γ bounded by a closed Jordan curve Γ. Let P be a point of Γ and let the cluster sets of $f(z)$ in γ and on Γ at P be s and S respectively. Then the maximum distance from any (finite) point O to a point of s is not greater than the maximum distance from O to a point of S.*

This theorem is proved by an application of the maximum principle for analytic functions. The essential reason for its power is that the bound of $f(z)$ does not enter into the conclusion. The theorem will now be shown to be equivalent to the following, stressed by Iversen in the second and third of his articles referred to above.

THEOREM 1'. *Let $f(z)$ be meromorphic in a domain γ bounded by a closed Jordan curve Γ. Let P be a point of Γ and let the cluster sets of $f(z)$ in γ and on Γ at P be s and S respectively. Then every frontier[5] point of s is a point of S.*

Theorem 1' implies Theorem 1. For if Q belonged to s and if OQ were greater than the distance from O to each point of S, there would evidently be a frontier point of s outside or on the circle of radius OQ about O and so not belonging to S.

Theorem 1 implies Theorem 1'. For if Theorem 1' were not true there would be a point β, a frontier point of s, but not belonging to S. Let every point of S be at distance greater than $d>0$ from β[6], and so at distance greater than $d/2$ from a point β' if

$$|\beta'-\beta|<d/2. \tag{1}$$

We can choose β' not belonging to s since β is a frontier point of s. Then $1/[f(z)-\beta']$ is bounded in a neighborhood of P, so by Theorem 1, taking O as the origin,

$$\frac{1}{|\beta'-\beta|} \leq \frac{2}{d} \tag{2}$$

which contradicts (1).

The theorem of Gross-Iversen will now be proved.

[5] A frontier point of a point set is a point every neighborhood of which contains a point both of the set and of its complement.

[6] If β is not finite, the argument proves the statement for the function $1/f(z)$ which is equivalent to proving it for $f(z)$.

THEOREM 2. *Let $f(z)$ be meromorphic in a domain γ bounded by a closed Jordan curve Γ. Let P be a point of Γ and let the cluster sets of $f(z)$ in γ and on Γ at P be s and S respectively. Then every point of s not belonging to S is assumed by $f(z)$ in γ in every neighborhood of P, with two possible exceptions. If there are any exceptions, they are convergence values of $f(z)$ in γ at P. If there are two exceptions, every other point of the (extended) plane is assumed by $f(z)$ in γ in every neighborhood of P. If α is a point of s but not of S, s includes the domain containing α and bounded only by points of S*[7].

(a) Every point of s not belonging to S and omitted by $f(z)$ in some neighborhood of P is a convergence value of $f(z)$ in γ at P. For suppose β is such a point. We suppose that β is finite; otherwise the function $1/f(z)$ would be substituted for $f(z)$. Then we can suppose also that $\beta = 0$. Since β does not belong to S, points A, B on Γ can be found so close to P, (different from P), one on each side of P, that every cluster value of $f(z)$ in γ at a point of APB other than P will be outside the circle $|w| = 2\varepsilon_1 > 0$, $w = f(z)$. The points A, B can be joined by a Jordan arc ACB (i. e. a one to one and continuous image of a line segment) lying within γ except for the points A, B, such that in the region R bounded by APB and by this arc, $f(z) \neq 0$, and such that $|f(z)| \geqq 2\varepsilon_2 > 0$ on ACB. Since $\beta = 0$ belongs to s, there is a point P_1 in R such that

$$|f(P_1)| < \varepsilon \tag{3}$$

where $\varepsilon > 0$, $\varepsilon \leqq \varepsilon_1$, $\varepsilon \leqq \varepsilon_2$. Let D_1 be the largest domain containing P_1 and consisting only of points at which $|f(z)| < \varepsilon$. At every boundary point of D_1 in γ, $|f(z)| = \varepsilon$. Then D_1 lies in R. The points A, B were so chosen that D_1 can have no boundary point on APB except possibly at P. Since $f(z) \neq 0$ in D_1, $|f(z)|$ cannot take on its minimum in D_1 at an interior point of D_1, applying the maximum principle to the function $1/f(z)$; (we suppose that the function $f(z)$ is not constant when we suppose that a point β exists). Then $f(z)$ must approach its lower bound in D_1, which is less than ε, on a sequence of points approaching a boundary point of D_1. Since $|f(z)| = \varepsilon$ at a boundary point of D_1 in R, the boundary point in question cannot be on the arc ACB. It was noted above that no boundary point of D_1 can be on the arc APB except perhaps at P. Then P must be a boundary point of D_1. Now the function $1/f(z)$ is not bounded in D_1. For suppose the contrary. Let P have the coördinate ξ and choose the constant $k \neq 0$ so that

[7] A domain will be said to be bounded only by points of a given point set S if every frontier point of the domain is a point of S.

$$|k(z-\xi)| \leqq 1 \tag{4}$$

for z in D_1. Then if σ is any positive number less than 1, the branches of $[k(z-a)]^\sigma$ form single-valued analytic functions in γ by the monodromic theorem. Moreover

$$\varphi(z) = \frac{[k(z-\xi)]^\sigma}{f(z)} \tag{5}$$

choosing some definite branch, is a bounded analytic function in D_1 which is not greater than $1/\varepsilon$ in modulus at all boundary points of D_1 except perhaps at P, $\lim_{z\to P} \varphi(z) = 0$ when z approaches P in D_1. Then $|\varphi(z)| \leqq 1/\varepsilon$ in D_1. Letting σ approach 0, we have

$$|f(z)| \geqq \varepsilon \tag{6}$$

in D_1 which contradicts (3). Thus $1/f(z)$ is not bounded in D_1,[8] and there is a point P_2 in D_1 such that $|f(P_2)| < \varepsilon/2$. Determine D_2 as the largest domain containing P_2 in which $|f(z)| < \varepsilon/2$. Then repeating the above argument, $D_2 \subset D_1$, D_2 has P as its only boundary point on Γ, and $1/f(z)$ is not bounded in D_2. This can be continued indefinitely, getting a sequence of domains $D_1, D_2, \cdots, D_1 \supset D_2 \supset D_3 \supset \cdots$ and a sequence of points $P_1, P_2, \cdots$ such that P_n belongs to D_n and such that

$$|f(z)| < \varepsilon/2^{n-1} \tag{7}$$

in D_n. Join P_n to P_{n+1} by a Jordan arc lying wholly in D_n. The curve composed of the sum of all these Jordan arcs is a continuous curve approaching P on which $f(z)$ has the limit $\beta = 0$. The curve must approach P as a limit since the sequence of domains $\{D_n\}$ shrinks to the point P. If the sequence of domains did not have this property, there would be a continuum common to them all, at the points of which $f(z) = 0$, which is impossible since $f(z)$ is not constant. It is easily seen that the curve can be replaced by a Jordan arc.

(b) The theorem can now be proved. If there were two points of s which did not belong to S and which were not assumed by $f(z)$ in some neighborhood of P, they would be convergence values of the function, as was just proved. Then by a well-known theorem of Lindelöf,[9] which the theorem of Gross-Iversen generalizes, $f(z)$ would assume every value in the extended plane in γ in every neighborhood of P with two possible exceptions, which must be the two exceptional values already known.

[8] This result could also have been proved using Seidel's theorem, but an examination of the boundary of D_1 would have been necessary.

[9] E. Lindelöf, Acta Societatis Scientiarum Fennicae, vol. 46 (1915), No. 4, p. 13.

If α is a point of s but not of S, s includes every point of the domain D containing α and bounded only by points of S. For otherwise there would be a frontier point of s in D and so not in S. This contradicts Theorem 1′.

It will be noted that the theorem does not state all the conclusions which can be drawn. For instance if there is just one exceptional value β, as defined in the theorem, let $f(z)$ converge to this value on the Jordan arc C. Then C divides the domain γ into two further domains, near P. Applying the theorem to these domains, we have the result that every point of one of the domains bounded only by points of the cluster sets of $f(z)$ on Γ at P on either side, and containing β is assumed by $f(z)$ in γ in every neighborhood of P, except the point β itself.

HARVARD UNIVERSITY,
January 13, 1932.

ON THE SUMMABILITY OF POWER SERIES[1].

By L. S. Bosanquet.

In this paper I obtain some extensions of the results of Hardy and Littlewood concerning the summability of power series analogous to those already obtained for Fourier series. The main results for Fourier series are embodied in the following two theorems[2].

Theorem A. *If $f(t)$ is integrable in the general Denjoy sense in the interval $(-\pi, \pi)$ and is periodic with period 2π, and if*

$$\varphi(t) = \tfrac{1}{2}\{f(x+t)+f(x-t)-2s\} \to 0\ (C, \alpha)^3$$

as $t \to 0$, where $\alpha \geqq 0$, then the Fourier series of $f(t)$ is summable $(C, \alpha+\delta)$ for $t = x$ to the sum s provided $\delta > 0$ and $\alpha + \delta \geqq 1$[4].

Theorem B. *If $f(t)$ is integrable in the general Denjoy sense in the interval $(-\pi, \pi)$ and is periodic with period 2π, and if its Fourier series is summable (C, α) to the sum s for $t = x$, where $\alpha \geqq -1$, then*

$$\varphi(t) = \tfrac{1}{2}\{f(x+t)+f(x-t)-2s\} \to 0\ (C, \alpha+1+\delta)$$

as $t \to 0$ provided $\delta > 0$.

There are analogous results for Allied series[5].

Hardy and Littlewood proved[6] that if[7] $f(z) = \sum a_n z^n$ *then the necessary and sufficient conditions that $\sum a_n$ be summable by some Cesàro mean to the sum A are that $a_n = O(n^K)$ for some K and that there be an integer k such that*

---

[1] Received February 4, 1932.

[2] Bosanquet (2). See also Hardy and Littlewood (9), (11) and (12), Paley (15), Verblunsky (16) and Wiener (18) and (19). For extensions to logarithmic scales of summability see Bosanquet and Linfoot (5) and (6).

[3] I. e. $\frac{\alpha}{t^\alpha}\int_0^t (t-u)^{\alpha-1}\varphi(u)\,du \to 0$, if $\alpha > 0$, and $\varphi(t) \to 0$ in the ordinary sense if $\alpha = 0$.

[4] The last restriction is unnecessary in the case of a function integrable in the Lebesgue sense. It arises from the fact that the nth Fourier coefficient is $o(n)$, and examples show that this cannot be improved, even in the case of a function integrable in the Cauchy-Lebesgue sense.

[5] Paley (15). See also Hardy and Littlewood (10) and Bosanquet and Linfoot (6).

[6] Hardy and Littlewood (9).

[7] Here and elsewhere $\sum$ is to be interpreted as $\sum_0^\infty$.

$$\frac{1}{1-z}\int_z^1 \frac{dz_1}{1-z_1}\int_{z_1}^1 \frac{dz_2}{1-z_2}\cdots\int_{z_{k-1}}^1 f(z_k)\,dz_k \to A$$

as $z\to 1$ *in* $|z|<1$.[8]

This condition may be expressed by saying that $f(z)\to A$ (H,k) as $z\to 1$ in $|z|<1$, and is equivalent[9] to

$$\frac{k}{(1-z)^k}\int_z^1 (u-z)^{k-1} f(u)\,du \to A$$

as $z\to 1$ in $|z|<1$, or $f(z)\to A\,(C,k)$ as $z\to 1$ in $|z|<1$. I show in this paper that if $a_n = o(n^\gamma)$, where $\gamma>0$, and $f(z)\to A\,(C,\alpha)$ as $z\to 1$ in $|z|<1$, where $\alpha\geqq 0$, then $\sum a_n$ is summable $(C,\alpha+\delta)$ to the sum A provided $\delta>0$ and $\alpha+\delta\geqq\gamma$.[10] Conversely, if $\sum a_n$ is summable (C,α) to the sum A, where $\alpha\geqq -1$, then $f(z)\to A\,(C,\alpha+1+\delta)$ as $z\to 1$ in $|z|\leqq 1$, provided $\delta>0$.[11] The case $\alpha=0$ of Theorem 1 was given by Dienes[12] for $\gamma\geqq 1$ and has recently been completed by Offord.[13]

Although the complications of real variable analysis are naturally absent from this problem there are some additional difficulties that arise, owing to the use of Cauchy's theorem, which seem to give Theorem 1, in particular, an independent interest. Here the argument is from the inside to the circumference of the circle of convergence, but in Theorem 2, where the argument is from the circumference to the inside, the difficulties are mainly formal. In the real variable problem the Riemann reduction to local hypotheses in Theorem A was easy,[14] but in the present problem I was at first unable to obtain the desired result in Theorem 1 without an additional hypothesis, such as $f(z)=o\{(1-|z|)^{-\gamma}\}$, which is stronger than $a_n=o(n^\gamma)$. This difficulty was finally resolved by Mr. A. C. Offord[15] and I am now able to quote from his work. The other difficulty is contained in the proof of Lemma 4,[16] which plays a fundamental rôle in the problem. In this connection I am indebted to Dr. Mary L. Cartwright for making the point clear to me.

---

[8] I. e. as z tends to 1 in any manner from within the unit circle. The limit thus exists *uniformly* in directions of approach.

[9] See lemma (4).

[10] Theorem 1.

[11] Theorem 2.

[12] Dienes (7). See also Dienes (8).

[13] Offord (14).

[14] See e. g. Bosanquet (2), 149.

[15] Offord (14).

[16] Due to Hardy and Littlewood.

1. Mean Continuity of Power Series.

Suppose $f(z) = \sum a_n z^n$, and let

$$f_\alpha(z) = \frac{1}{\Gamma(\alpha)} \int_z^1 (u-z)^{\alpha-1} f(u)\, du, \qquad \alpha > 0,$$

the integral being taken along the straight line $(z, 1)$.[17] Then $f_\alpha(z)$ is called the fractional integral of $f(z)$ of order α taken at $z = 1$. Writing $f_0(z) = f(z)$, we have for $k \geqq 1$,

$$f_k(z) = \int_z^1 f_{k-1}(u)\, du.$$

If

$$(1-z)^{-\alpha} \Gamma(\alpha+1) f_\alpha(z) \to A$$

as $z \to 1$ in $|z| < 1$, where $\alpha \geqq 0$, we say that $f(z) \to A\,(C, \alpha)$ as $z \to 1$ in $|z| < 1$. We also define the corresponding (H, k) limit in the manner already indicated.[18]

It will be seen that if $f(z) \to A\,(C, \alpha)$ as $z \to 1$ in $|z| < 1$ then $f(z) \to A\,(C, 0)$ as $z \to 1$ in any angle inside the unit circle whose vertex is $z = 1$, and so we say that $f(z)$ is continuous (C, α) at $z = 1$ from within the unit circle.

The properties of these fractional integrals and means are embodied in the next three lemmas.

Lemma 1. *If $f_\alpha(z)$ exists, where $\alpha > 0$ and $|z| < 1$, then*

$$f_{\alpha+\beta}(z) = \frac{1}{\Gamma(\beta)} \int_z^1 (u-z)^{\beta-1} f_\alpha(u)\, du,$$

provided $\beta > 0$.

Lemma 2. *If $f(z) \to A\,(C, \alpha)$ as $z \to 1$ in $|z| < 1$, where $\alpha \geqq 0$, then $f(z) \to A\,(C, \alpha+\beta)$ provided $\beta \geqq 0$.*

Lemma 3. *If k is an integer and $f(z) \to A\,(C, k)$ as $z \to 1$ in $|z| < 1$, then $f(z) \to A\,(H, k)$ as $z \to 1$ in $|z| < 1$, and conversely.*

The proofs of these lemmas are similar to their real variable analogues and are discussed elsewhere.[19]

2.

Theorem 1. *If $f(z) = \sum a_n z^n$, where $a_n = o(n^\gamma)$, $\gamma > 0$, and $f(z) \to A\,(C, \alpha)$ as $z \to 1$ in $|z| < 1$, where $\alpha \geqq 0$, then $\sum a_n$ is summable (C, β) to the sum A, provided $\beta > \alpha$ and $\beta \geqq \gamma$.*

The proof depends upon the following five lemmas.

---

[17] Here and elsewhere $\int_z^1$ is to be interpreted as $\lim_{\zeta \to 1} \int_z^\zeta$, where $\zeta \to 1$ along the line $(z, 1)$.

[18] p. 759.

[19] See Bosanquet and Cartwright, 4. Cf. Hardy and Littlewood, 13, and Bosanquet, 1 and 3.

LEMMA 4. *If $f(z)$ satisfies the hypotheses of Theorem* 1 *and k is a positive integer, then $f_k(z)$ is regular for $|z|<1$ and*

$$f_k(z) = \int_z^1 f_{k-1}(u)\,du = \int_0^1 f_{k-1}(u)\,du - \int_0^z f_{k-1}(u)\,du,$$

the integrals being taken along straight lines.

This is analogous to a lemma given by Hardy and Littlewood.[20] It is important to notice that its truth depends (a) upon the existence of $f_k(z)$ as a Cauchy integral, which is a consequence of the second hypothesis, and (b) upon the fact that $f(z) = O\{(1-|z|)^{-K}\}$ in an angle inside the unit circle whose vertex is $z=1$, which is a consequence of the first hypothesis. The proof involves difficult theorems of Phragmén and Lindelöf.

LEMMA 5. *Under the hypotheses of Theorem* 1 *$f_\lambda(z)$ is regular in $|z|<1$ for every $\lambda \geqq 0$.*

This lemma is proved elsewhere.[21] It may also be deduced from Hardy and Littlewood's theorem.

LEMMA 6. *Under the hypotheses of Theorem* 1, *if k is a positive integer less than $\gamma+1$, then*

$$f_k(z) = o\{(1-|z|)^{k-\gamma-1}\}$$

as $|z|\to 1$ from within the unit circle.

By Lemma 4, we have

$$f_1(z) = C - \sum \frac{a_n}{n+1} z^{n+1} = C + \sum o(n^{\gamma-1})\,|z|^{n+1} = o\{(1-|z|)^{-\gamma}\},$$

and the complete result is readily established by k applications of this reasoning.

LEMMA 7. *If $f(z) = \sum a_n z^n$ is regular for $|z|<1$ and $f(z)\to 0\,(C,\alpha)$ as $z\to 1$ in $|z|<1$, and if h is the greatest integer not greater than α, then*

$$f_{h+1}(z) = o(|1-z|^{h+1})$$

as $z\to 1$ in $|z|<1$.

This is simply a particular case of Lemma 2.

LEMMA 8. *Under the hypotheses of Theorem* 1, *with $A=0$, if $0<\xi<\eta \leqq 2\xi$, and h is the greatest integer not greater than α, then*

$$f_{h+1}\{(1-\varrho)\,e^{i\eta}\} - f_{h+1}\{(1-\varrho)\,e^{i\xi}\} = o\{(\xi+\varrho)^\alpha\,(\eta-\xi)^{1+h-\alpha}\}$$

as ξ and $\varrho\to 0$.

[20] Hardy and Littlewood (9), 80, though they attribute the proof to A. E. Ingham. See also Bosanquet and Cartwright, (4), where more general results are given.

If for a moment we assume the theorem of Hardy and Littlewood stated in the introduction it follows, under the hypotheses, from Lemmas 2 and 3 and their theorem that $\sum a_n$ is summable (C). It then follows from Hölder's extension of Frobenius' theorem that $f(z)$ is continuous in the closed triangle $(0, 1, z)$ and Lemma 4 simply becomes Cauchy's theorem for a triangle. But of course this is not a proof since Lemma 4 is used in the proof of Hardy and Littlewood's theorem.

[21] See Bosanquet and Cartwright, 4.

Let $r = 1-\rho$. Then, by Lemmas 1 and 5 and Cauchy's theorem, we have

$$
\begin{aligned}
&f_{h+1}(re^{i\eta}) - f_{h+1}(re^{i\xi}) \\
&\quad = \frac{1}{\Gamma(1+h-\alpha)} \int_{re^{i\xi}}^{1} (u - re^{i\xi})^{h-\alpha} f_\alpha(u)\, du \\
&\qquad - \frac{1}{\Gamma(1+h-\alpha)} \int_{re^{i\eta}}^{1} (u - re^{i\eta})^{h-\alpha} f_\alpha(u)\, du \\
&\quad = \frac{1}{\Gamma(1+h-\alpha)} \int_{re^{i\xi}}^{re^{i\eta}} (u - re^{i\eta})^{h-\alpha} f_\alpha(u)\, du \\
&\qquad + \frac{1}{\Gamma(1+h-\alpha)} \int_{re^{i\xi}}^{1} \{(u - re^{i\xi})^{h-\alpha} - (u - re^{i\eta})^{h-\alpha}\} f_\alpha(u)\, du,
\end{aligned}
$$

the integrals being taken along straight lines. Here, in addition to Lemma 5, we have used the fact that $f_\alpha(z)$ is continuous at $z = 1$ from within the unit circle, and also that $\alpha - h < 1$.

Observing that $|1 - re^{i\theta}| = O(\rho + \theta)$ as ρ and $\theta \to 0$, we see that in each integral $f_\alpha(u) = o\{(\xi+\rho)^\alpha\}$.

The length of the path of the first integral is $O(\eta - \xi)$, and hence this integral is easily seen to be

$$o\{(\xi+\rho)^\alpha (\eta-\xi)^{1+h-\alpha}\}.$$

The second integral is

$$o\left\{(\xi+\rho)^\alpha \int_{re^{i\xi}}^{1} |(u - re^{i\xi})^{h-\alpha} - (u - re^{i\eta})^{h-\alpha}|\, |du|\right\},$$

the integral being taken along the straight line $(re^{i\xi}, 1)$. If we now write $|u - re^{i\xi}| = v$, $|1 - re^{i\xi}| = \sigma$, this is

$$
\begin{aligned}
&o\left[(\xi+\rho)^\alpha \int_0^\sigma \{v^{h-\alpha} - (v + \eta - \xi)^{h-\alpha}\}\, dv\right] \\
&\quad = o\left[(\xi+\rho)^\alpha \{\sigma^{1+h-\alpha} - (\sigma+\eta-\xi)^{1+h-\alpha} + (\eta-\xi)^{1+h-\alpha}\}\right] \\
&\quad = o\left[(\xi+\rho)^\alpha (\eta-\xi)^{1+h-\alpha}\right],
\end{aligned}
$$

since $\sigma = O(\xi+\rho)$ and $x^k - y^k = O\{(x-y)^k\}$ when $0 < k \leqq 1$, $x > y > 0$ and $x \to 0$.

Proof of Theorem 1. There is no loss of generality in supposing that $A = 0$ and $\beta \geqq \gamma$. Writing

$$
\begin{aligned}
\sum S_n^\beta z^n &= (1-z)^{-1-\beta} \sum a_n z^n, \\
\sum C_n^{n+\beta} z^n &= (1-z)^{-1-\beta},
\end{aligned}
$$

we have

$$\frac{S_n^\beta}{C_n^{n+\beta}} = O(n^{-\beta}) \int_{|z|=r} f(z) \frac{dz}{(1-z)^{1+\beta} z^{n+1}},$$

where $0 < r < 1$, and we have to show that $S_n^\beta / C_n^{n+\beta} = o(1)$.

Now it has been shown by Offord[22], on the sole hypothesis $a_n = o(n^\gamma)$, that this may be reduced to showing that

$$\int_{am\, z=-\delta}^{\delta} f(z)\, Q(z)\, z^{-n-1}\, dz = o(n^\beta),$$

where the integral is taken along $|z| = 1-1/n$, $Q(z) = (1-z)^{-1-\beta}+c_0+c_1 z$ and c_0, c_1 are constants depending on δ, which is arbitrary but fixed. It will be seen that the constants c_0, c_1 play no further rôle in the proof. This is due to the application of Lemma 6, which depends on Lemma 4, and so ultimately is due to the local hypothesis.

After h integrations by parts, where h is the greatest integer not greater than α, we obtain

$$\begin{aligned}
&\int_{am\, z=-\delta}^{\delta} f(z)\, Q(z)\, z^{-n-1}\, dz \\
&\quad = \left[-f_1(z)\, Q(z)\, z^{-n-1}\right]_{am\, z=-\delta}^{\delta} + \int_{am\, z=-\delta}^{\delta} f_1(z)\, \frac{d}{dz}\{Q(z)\, z^{-n-1}\}\, dz \\
&\quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&\quad = \sum_{k=1}^{h}\left[-f_k(z)\, \frac{d^{k-1}}{dz^{k-1}}\{Q(z)\, z^{-n-1}\}\right]_{am\, z=-\delta}^{\delta} \\
&\qquad\qquad + \int_{am\, z=-\delta}^{\delta} f_h(z)\, \frac{d^h}{dz^h}\{Q(z)\, z^{-n-1}\}\, dz,
\end{aligned}$$

the integrals being taken along $|z| = 1-1/n$.

The terms outside the integral are

$$\sum_{k=1}^{h} o(n^{-k+\gamma+1})\, O(n^{k-1}) = o(n^\gamma)$$

by Lemma 6. Now write

$$\begin{aligned}
J(n) &= \int_{am\, z=0}^{\delta} f_h(z)\, \frac{d^h}{dz^h}\{Q(z)\, z^{-n-1}\}\, dz \\
&= \int_{am\, z=0}^{1/n} + \sum_{\nu=1}^{\tau-1}\int_{am\, z=\nu/n}^{(\nu+1)/n} + \int_{am\, z=\tau/n}^{\delta} \\
&= J_0 + \sum_{\nu=1}^{\tau-1} J_\nu + J_\tau,
\end{aligned}$$

where τ is the greatest integer not greater than $n\delta$. Then we need only show that $J(n) = o(n^\beta)$, the proof for the remaining integral being similar.

Let δ be chosen so that, for $0 < \xi < \eta \leqq 2\xi \leqq \delta$ and $n > 1/\delta$,

$$|f_{h+1}(re^{i\eta}) - f_{h+1}(re^{i\xi})| < \varepsilon\left(\xi + \frac{1}{n}\right)^\alpha (\eta-\xi)^{1+h-\alpha},$$

where $r > 1/n$ and ε is an arbitrary positive number. Then, for $0 < \nu \leqq \tau-1$,

[22] Offord (14).

$$J_\nu = \int_{am\, z=\nu/n}^{(\nu+1)/n} f_h(z) \frac{d^h}{dz^h} \{Q(z)\, z^{-n-1}\}\, dz$$

$$= \left[-\{f_{h+1}(z) - f_{h+1}(re^{i\nu/n})\} \frac{d^h}{dz^h} \{Q(z)\, z^{-n-1}\} \right]_{am\, z=\nu/n}^{(\nu+1)/n}$$

$$+ \int_{am\, z=\nu/n}^{(\nu+1)/n} \{f_{h+1}(z) - f_{h+1}(re^{i\nu/n})\} \frac{d^{h+1}}{dz^{h+1}} \{Q(z)\, z^{-n-1}\}\, dz,$$

and hence

$$|J_\nu| < \varepsilon \left(\frac{\nu+1}{n}\right)^\alpha \left(\frac{1}{n}\right)^{1+h-\alpha} \sum_{\lambda=0}^{h} O\left\{\left(\frac{\nu}{n}\right)^{-\beta-1-\lambda} n^{h-\lambda}\right\}$$

$$+ \varepsilon \left(\frac{\nu+1}{n}\right)^\alpha \left(\frac{1}{n}\right)^{1+h-\alpha} \frac{1}{n} \sum_{\lambda=0}^{h+1} O\left\{\left(\frac{\nu}{n}\right)^{-\beta-1-\lambda} n^{h+1-\lambda}\right\}$$

$$< \varepsilon\, \nu^{\alpha-\beta-1}\, O(n^\beta)$$

uniformly in ν. In the same way we can show that

$$|J_\tau| < \varepsilon\, \tau^{\alpha-\beta-1}\, O(n^\beta).$$

Thus we have

$$\sum_{\nu=1}^{\tau} |J_\nu| < \varepsilon\, O(n^\beta) \sum_{\nu=1}^{\tau} \nu^{\alpha-\beta-1} = \varepsilon\, O(n^\beta)\, O(1),$$

since $\beta > \alpha$.

Finally we have

$$J_0 = \int_{am\, z=0}^{1/n} f_h(z) \frac{d^h}{dz^h} \{Q(z)\, z^{-n-1}\}\, dz$$

$$= \left[-f_{h+1}(z) \frac{d^h}{dz^h} \{Q(z)\, z^{-n-1}\} \right]_{am\, z=0}^{1/n}$$

$$+ \int_{am\, z=0}^{1/n} f_{h+1}(z) \frac{d^{h+1}}{dz^{h+1}} \{Q(z)\, z^{-n-1}\}\, dz$$

$$= o\left\{\left(\frac{1}{n}\right)^{h+1}\right\} \sum_{\lambda=0}^{h} O\left\{\left(\frac{1}{n}\right)^{-\beta-1-\lambda} n^{h-\lambda}\right\}$$

$$+ o\left\{\left(\frac{1}{n}\right)^{h+1}\right\} \frac{1}{n} \sum_{\lambda=0}^{h+1} O\left\{\left(\frac{1}{n}\right)^{-\beta-1-\lambda} n^{h+1-\lambda}\right\}$$

$$= o(n^\beta) + o(n^\beta).$$

Thus $J(n) = o(n^\beta)$ and the proof is complete.

3.

In this section I consider the reciprocal problem. For this purpose it is necessary to construct a function to play the rôle of Young's function[23] in the corresponding real variable problem.[24]

[23] Or, more precisely, the function $\gamma_\beta(t) = \dfrac{\Gamma(\beta)\, C_\beta(t)}{t^\beta} = \displaystyle\int_0^1 (1-u)^{\beta-1} \cos tu\, du$, where $C_\beta(t)$ is Young's function. Young (17).

[24] Cf. Bosanquet and Linfoot (5) and (6) where a similar process is carried out.

The function $G_\beta(n, z)$. Let

$$G_\beta(n, z) = \frac{1}{(1-z)^\beta} \int_z^1 (u-z)^{\beta-1} u^n \, du,$$

where $\beta > 0$, n is a positive integer and $|z| \leqq 1$. The integral is as usual taken along the straight line $(z, 1)$.

Let

$$\begin{aligned} \Delta G_\beta(n, z) &= G_\beta(n, z) - G_\beta(n+1, z) \\ &= \frac{1}{(1-z)^\beta} \int_z^1 (u-z)^{\beta-1} (1-u) u^n \, du. \end{aligned}$$

In general let

$$\begin{aligned} \Delta^k G_\beta(n, z) &= \Delta^{k-1} G_\beta(n, z) - \Delta^{k-1} G_\beta(n+1, z) \\ &= \frac{1}{(1-z)^\beta} \int_z^1 (u-z)^{\beta-1} (1-u)_k \, u^n \, du, \end{aligned}$$

where $\Delta^0 G_\beta(n, z) = G_\beta(n, z)$.

Then the properties of the function $G_\beta(n, z)$ and its differences are embodied in the following lemma.

LEMMA 9. *If* $\beta > 0$, $|z| \leqq 1$ *and* k *is a non-negative integer, then*

$$|\Delta^k G_\beta(n, z)| \begin{cases} \leqq A|1-z|^k, & \text{for all } n \geqq 0, \\ \leqq A|1-z|^{k-\beta} n^{-\beta}, & 0 < \beta \leqq k+1, \\ \leqq A|1-z|^{-1} n^{-1-\kappa}, & \beta > k+1, \end{cases}$$

where A *denotes some number independent of* n *and* z.

Let $1-z = (1-y)e^{i\alpha}$ and $1-u = (1-v)e^{i\alpha}$. Then

$$\begin{aligned} \Delta^k G_\beta(n, z) &= \left| \frac{1}{(1-z)^\beta} \int_z^1 (u-z)^{\beta-1} (1-u)^k u^n \, du \right| \\ &\leqq \frac{1}{|1-z|^\beta} \int_z^1 |u-z|^{\beta-1} |1-u|^k \, |du| \\ &= \frac{1}{(1-y)^\beta} \int_y^1 (v-y)^{\beta-1} (1-v)^k \, dv \\ &\leqq (1-y)^{k-\beta} \int_y^1 (v-y)^{\beta-1} \, dv \\ &\leqq A(1-y)^k, \end{aligned}$$

which proves the first part.

Now write

$$\begin{aligned} \frac{1}{(1-z)^\beta} \int_z^1 (u-z)^{\beta-1} (1-u)^k u^n \, du &= \frac{1}{(1-z)^\beta} \int_z^{z+\frac{1}{n}e^{i\alpha}} + \frac{1}{(1-z)^\beta} \int_{z+\frac{1}{n}e^{i\alpha}}^1 \\ &= I_1 + I_2. \end{aligned}$$

Then, if $\beta \leqq 1$,

$$\begin{aligned} |J_1| &\leqq \frac{1}{(1-y)^\beta} \int_y^{y+\frac{1}{n}} (v-y)^{\beta-1}(1-v)^k\, dv \\ &\leqq (1-y)^{k-\beta} \int_y^{y+\frac{1}{n}} (v-y)^{\beta-1}\, dv \\ &\leqq A(1-y)^{k-\beta} n^{-\beta}, \end{aligned}$$

while

$$\begin{aligned} |J_2| &= \frac{1}{(1-y)^\beta} \left| \int_{y+\frac{1}{n}}^1 (v-y)^{\beta-1}(1-v)^k u^n\, dv \right| \\ &\leqq (1-y)^{k-\beta} \left(\frac{1}{n}\right)^{\beta-1} 2 \max_{y \leqq \lambda < \mu \leqq 1} \left| \int_\lambda^\mu u^n\, dv \right| \\ &\leqq A(1-y)^{k-\beta} n^{-\beta}. \end{aligned}$$

Here we have used the fact that $(v-y)^{\beta-1}(1-v)^k$ is a positive decreasing function, and applied the mean value theorem separately to the real and imaginary parts of u^n.

The lemma is thus proved for the case $\beta \leqq 1$. It can be completed for the case $\beta > 1$ by means of a preliminary integration by parts, and reasoning of a similar nature.

THEOREM 2. *If $\sum a_n$ is summable (C, α) to the sum A, where $\alpha \geqq -1$, then $f(z) = \sum a_n z^n$ is regular for $|z| < 1$ and $f(z) \to A$ (C, β) as $z \to 1$ in $|z| \leqq 1$, provided $\beta > \alpha + 1$.*

The proof depends on the following two lemmas, analogous to Lemmas 7 and 8.

LEMMA 10. *If $\sum a_n$ is summable (C, α) to the sum 0, where $\alpha \geqq -1$, then*

$$S_n^{h+1} = o(n^{h+1}).$$

This is simply a consequence of the consistency theorem for Cesàro summation.

LEMMA 11. *If $\sum a_n$ is summable (C, α) to the sum 0, where $\alpha > -1$, and $0 < p < q \leqq 2p$, where p and q are integers, then*

$$S_q^{h+1} - S_p^{h+1} = o\{p^\alpha (q-p)^{1+h-\alpha}\}$$

as $p \to \infty$.

We have

$$\begin{aligned} &S_q^{h+1} - S_p^{h+1} \\ &= \frac{1}{\Gamma(1+h-\alpha)} \sum_{n=0}^{q} S_n^\alpha \frac{\Gamma(q-n+1+h-\alpha)}{\Gamma(q-n+1)} \\ &\quad - \frac{1}{\Gamma(1+h-\alpha)} \sum_{n=0}^{p} S_n^\alpha \frac{\Gamma(p-n+1+h-\alpha)}{\Gamma(p-n+1)} \end{aligned}$$

$$= \frac{1}{\Gamma(1+h-\alpha)} \sum_{n=p+1}^{q} S_n^{\alpha} \frac{\Gamma(q-n+1+h-\alpha)}{\Gamma(q-n+1)}$$

$$- \frac{1}{\Gamma(1+h-\alpha)} \sum_{n=0}^{p} S_n^{\alpha} \left\{ \frac{\Gamma(p-n+1+h-\alpha)}{\Gamma(p-n+1)} - \frac{\Gamma(q-n+1+h-\alpha)}{\Gamma(q-n+1)} \right\}.$$

By Stirling's formula, the first term is

$$o\left\{ p^{\alpha} \sum_{n=p+1}^{q} (q-n+1)^{h-\alpha} \right\} = o\{p^{\alpha}(q-p)^{1+h-\alpha}\},$$

while, if $\alpha \geqq 0$, the second is

$$o\left[p^{\alpha} \sum_{n=0}^{p} \{(p-n+1)^{h-\alpha} - (q-n+1)^{h-\alpha}\} \right]$$
$$= o[p^{\alpha}\{p^{1+h-\alpha} - q^{1+h-\alpha} + (q-p)^{1+h-\alpha}\}]$$
$$= o\{p^{\alpha}(q-p)^{1+h-\alpha}\},$$

as in Lemma 8, since $0 < 1+h-\alpha \leqq 1$.

If $-1 < \alpha < 0$ the last result is still true. The details are left to the reader.

Proof of Theorem 2. There is no loss of generality in supposing that $A = 0$, $\alpha > -1$ and $\beta < h+2$, where h is the greatest integer not greater than α. Since by hypothesis $\sum a_n$ is summable (C) to the sum 0, it follows that $f(z)$ is continuous in any angle $|am(1-z)| \leqq \frac{1}{2}\pi - \delta$ and tends to zero as $z \to 1$ in the angle. Therefore, if $|z| \leqq 1$,

$$\frac{1}{(1-z)^{\beta}} \int_z^1 (u-z)^{\beta-1} f(u)\, du$$
$$= \frac{1}{(1-z)^{\beta}} \int_z^1 (u-z)^{\beta-1} \sum a_n u^n\, du$$
$$= \frac{1}{(1-z)^{\beta}} \int_z^1 (u-z)^{\beta-1} \lim_{r \to 1-0} \sum a_n r^n u^n\, du$$
$$= \lim_{r \to 1-0} \frac{1}{(1-z)^{\beta}} \int_z^1 (u-z)^{\beta-1} \sum a_n r^n u^n\, du$$
$$= \lim_{r \to 1-0} \sum a_n r^n \frac{1}{(1-z)^{\beta}} \int_z^1 (u-z)^{\beta-1} u^n\, du$$
$$= \lim_{r \to 1-0} \sum a_n r^n G_{\beta}(n, z)$$
$$= \sum_{(A)} a_n G_{\beta}(n, z).$$

Now write

$$S^0_{N,r} = \sum_{n \leqq N} a_n r^n, \qquad S^k_{N,r} = \sum_{n \leqq N} S^{k-1}_{n,r}.$$

Then, for a fixed positive $r<1$, $a_n r^n = o(1)$, and hence, if k be any positive integer, $S^k_{N,r} = o(N^{k+1})$.

Therefore, for a fixed z, $|z|<1$, and a fixed positive $r<1$, we have by $h+1$ partial summations,

$$\begin{aligned}
&\sum_{n \leqq N} a_n r^n G_\beta(n, z) \\
&\quad = S^0_{N,r} G_\beta(N+1, z) + \sum_{n \leqq N} S^0_{n,r} \Delta G_\beta(n, z) \\
&\quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&\quad = \sum_{k=0}^{h} S^k_{N,r} \Delta^k G_\beta(N+1, z) + \sum_{n \leqq N} S^h_{n,r} \Delta^{h+1} G_\beta(n, z) \\
&\quad = \sum_{k=0}^{h} o(N^{k+1}) O(N^{-k-1}) + \sum_{n \leqq N} S^h_{n,r} \Delta^{h+1} G_\beta(n, z).
\end{aligned}$$

Now letting $N \to \infty$, we have

$$\sum a_n r^n G_\beta(n, z) = \sum S^h_{n,r} \Delta^{h+1} G_\beta(n, z),$$

and then, letting $r \to 1$, we obtain

$$\sum_{(A)} a_n G_\beta(n, z) = \sum S^h_n \Delta^{h+1} G_\beta(n, z).$$

It remains therefore to show that

$$S(z) = \sum S^h_n \Delta^{h+1} G_\beta(n, z) = o(1)$$

as $z \to 1$ in $|z| \leqq 1$.

Let ω be the greatest integer not greater than $|1-z|^{-1}$, and write

$$\sum S^h_n \Delta^{h+1} G_\beta(n, z) = \sum_{n=0}^{\omega} + \sum_{\nu=1}^{\infty} \sum_{n=\nu\omega+1}^{(\nu+1)\omega} = J_0 + \sum_{\nu=1}^{\infty} J_\nu.$$

Then

$$\begin{aligned}
J_0 &= \sum_{n=0}^{\omega} S^h_n \Delta^{h+1} G_\beta(n, z) \\
&= S^{h+1}_\omega \Delta^{h+1} G_\beta(\omega+1, z) + \sum_{n=0}^{\omega} S^{h+1}_n \Delta^{h+2} G_\beta(n, z) \\
&= o(\omega^{h+1}) O(|1-z|^{h+1}) + \sum_{n=0}^{\omega} o(n^{h+1}) O(|1-z|^{h+2}) \\
&= o(1) + o(1).
\end{aligned}$$

Again, by a partial summation, if $\nu > 0$,

$$\begin{aligned}
J_\nu &= \sum_{n=\nu\omega+1}^{(\nu+1)\omega} S_n^h \, \Delta^{h+1} G_\beta(n, z) \\
&= \{S_{(\nu+1)\omega}^{h+1} - S_{\nu\omega}^{h+1}\} \, \Delta^{h+1} G_\beta\{(\nu+1)\omega + 1, z\} \\
&\quad + \sum_{n=\nu\omega}^{(\nu+1)\omega} \{S_n^{h+1} - S_{\nu\omega}^{h+1}\} \, \Delta^{h+2} G_\beta(n, z) \\
&= o\{(\nu\omega)^\alpha \, \omega^{1+h-\alpha}\} \, O\{|1-z|^{1+h-\beta} (\nu\omega)^{-\beta}\} \\
&\quad + \sum_{n=\nu\omega}^{(\nu+1)\omega} o\{(\nu\omega)^\alpha (n-\nu\omega)^{1+h-\alpha}\} \, O\{|1-z|^{2+h-\beta} n^{-\beta}\} \\
&= o\{\nu^{\alpha-\beta} \, \omega^{1+h-\beta} \, |1-z|^{1+h-\beta}\} \\
&\quad + o(\nu\omega)^\alpha \, O|1-z|^{2+h-\beta} \sum_{n=\nu\omega}^{(\nu+1)\omega} (n-\nu\omega)^{1+h-\alpha} \, n^{-\beta} \\
&= o(\nu^{\alpha-\beta}) + o(\nu\omega)^\alpha \, O(|1-z|^{2+h-\beta}) \, O\{\omega^{2+h-\alpha} (\nu\omega)^{-\beta}\} \\
&= o(\nu^{\alpha-\beta}) + o(\nu^{\alpha-\beta})
\end{aligned}$$

uniformly in ν as $z \to 1$ in $|z| \leqq 1$. Hence

$$\sum_{\nu=1}^{\infty} J_\nu = \sum_{\nu=1}^{\infty} o(\nu^{\alpha-\beta}) = o(1) \sum_{\nu=1}^{\infty} \nu^{\alpha-\beta} = o(1) \, O(1),$$

since $\beta > \alpha + 1$. This completes the proof.

Finally I state a theorem which is an extension of Theorem 1.

THEOREM 3. *If* $f(z) = \sum a_n z^n$, *where* $a_n = o(n^\gamma)$, $\gamma > 0$, *and if* $f(z) \to A$ (C, α) *as* $z \to 1$ *in* $|1-z| \leqq B(1-|z|)^\lambda$, $0 < \lambda < 1$, *and further, in the neighbourhood of* $z = 1$, $f(z) = o\{(1-|z|)^{-k}\}$ *as* $|z| \to 1$, $k > 0$, *then* $\sum a_n$ *is summable* (C, β) *to the sum* A, *provided* $\beta > \alpha$, $\beta \geqq \gamma$ *and* $\beta \geqq k/(1-\lambda)$.

This theorem was originally obtained by Miss Cartwright in a rather less precise form and was the starting point of our joint work.[25] Theorem 1 corresponds to the case $\lambda = 0$.

LIST OF REFERENCES.

1. L. S. Bosanquet. "On Abel's integral equation and fractional integrals." *Proc. London Math. Soc.* (2) 31 (1930), pp. 134–143.
2. — "On the summability of Fourier series." *Proc. London Math. Soc.* (2) 31 (1930), pp. 144–164.
3. — "Note on the limit of a function at a point." *Journal London Math. Soc.* 7 (1932), pp. 100–105.
4. L. S. Bosanquet and M. L. Cartwright. "On the Hölder and Cesàro means of an analytic function." Unpublished.

[25] Cf. Bosanquet and Cartwright, 4, Theorem A. The theorem given above is the result of combining Theorem 1a of our paper with Theorem 1 of this paper.

5. L. S. Bosanquet and E. H. Linfoot. "On the zero order summability of Fourier series." *Journal London Math. Soc.* 6 (1931), pp. 117–126.
6. — "Generalized means and the summability of Fourier series." *Quarterly Journal* (Oxford Series) 2 (1931), pp. 207–229.
7. P. Dienes. "Sur la sommabilité de la série de Taylor." *Comptes Rendus* 153 (1911), pp. 802–5.
8. — "The Taylor series." Oxford (1931).
9. G. H. Hardy and J. E. Littlewood. "Solution of the Cesàro summability problem for power series and Fourier series." *Math. Zeit.* 19 (1924), pp. 67—96.
10. — "The allied series of a Fourier series." *Proc. London Math. Soc.* (2) 24 (1926), pp. 211–246.
11. — "Notes on the theory of series III: on the summability of the Fourier series of a nearly continuous function." *Proc. Cambridge. Phil. Soc.* 23 (1927), pp. 681–684.
12. — "Notes on the theory of series VII: on Young's convergence criterion for Fourier Series." *Proc. London Math. Soc.* (2) 28 (1928), pp. 301–311.
13. — "The equivalence of certain integral means." *Proc. London. Math. Soc. Records.* (2) 22 (1924), pp. xl–xliii.
14. A. C. Offord. "On the summability of power series." *Proc. London Math. Soc.* (2) 33 (1932), pp. 467–480.
15. R. E. A. C. Paley. "On the Cesàro summability of Fourier series and Allied series." *Proc. Cambridge Phil. Soc.* 26 (1930), pp. 173–203.
16. S. Verblunsky. "Note on the sum of oscillating series." *Proc. Cambridge Phil. Soc.* 26 (1930), pp. 152–157.
17. W. H. Young. "On infinite integrals involving a generalization of the sine and cosine functions." *Quart. J. of Math.* 43 (1912), pp. 161–177.
18. N. Wiener. "A type of Tauberian theorem applying to Fourier series." *Proc. London Math. Soc.* (2) 30 (1929), pp. 1–8.
19. — "Tauberian theorems." *Annals of Math.* 33 (1932), pp. 1–100.

NON-SUMMABILITY OF THE CONJUGATE SERIES OF A FOURIER SERIES.[1]

By B. N. Prasad.

1. There is a very intimate relation between the conjugate function

$$g(\theta) = \frac{1}{2\pi}\int_0^{\pi} \{f(\theta+t)-f(\theta-t)\}\cot\frac{1}{2}t\,dt \tag{1.1}$$

in which the integral is taken in the sense of Cauchy, and the conjugate series

$$\sum_{n=1}^{\infty}(b_n\cos n\theta - a_n\sin n\theta) \tag{1.2}$$

of the Fourier series corresponding to the function $f(\theta)$ which is periodic and integrable in the sense of Lebesgue. The various properties of the conjugate series regarding convergence, summability, etc. are made to depend upon the behaviour of the integral (1.1). Thus, if (1.1) exists as an ordinary convergent Cauchy integral, it is known that the series (1.2) is summable (C, δ) for every $\delta > 1$,[2] and summable (C, δ) for every $\delta > 0$ if $f(\theta)$ is bounded,[3] while if the oscillating integral (1.1) is summable in some Cesàro sense, the conjugate series is also summable (C, δ) for a suitable value of δ.[4] The question arises: what is the behaviour of the conjugate series when the integral (1.1) diverges to $+\infty$ or $-\infty$? I formally prove that in that case the Abel limit of the conjugate series also diverges and hence *a fortiori* the series cannot be summable (C, δ) for any positive value of δ.

2. Theorem. *If the integral*

$$g(\theta) = \frac{1}{2\pi}\int_0^{\pi} \{f(\theta+t)-f(\theta-t)\}\cot\frac{1}{2}t\,dt$$

diverges to $+\infty$ *or* $-\infty$, *the Abel limit of the conjugate series* (1.2) *will also diverge to the same value.*

Let us write

$$\psi(t) = f(\theta+t)-f(\theta-t),$$

and

$$\chi(t) = \int_t^{\pi} \psi(t)\cot\frac{1}{2}t\,dt.$$

We have

$$F(x) = \pi\sum_{n=1}^{\infty}(b_n\cos n\theta - a_n\sin n\theta)\,x^n$$

$$= \frac{1}{2}\int_{-\pi}^{\pi} f(\alpha)\frac{2x\sin(\alpha-\theta)}{1-2x\cos(\alpha-\theta)+x^2}\,d\alpha$$

[1] Received February 19, 1932.

[2] Paley, **2** (Theorem 2, with $\alpha = 1$).

[3] Prasad, **3**, p. 274.

[4] Hardy and Littlewood, **1** (Theorem 3); Paley, *ibid.* (Theorem 2).

$$= \int_0^\pi \{f(\theta+t)-f(\theta-t)\}\frac{x\sin t}{1-2x\cos t+x^2}\,dt$$

$$= \int_0^\pi \psi(t)\cot\frac{t}{2}\,\frac{2x\sin^2\frac{t}{2}}{1-2x\cos t+x^2}\,dt.$$

Integrating by parts, we get

$$F(x) = \left[-\chi(t)\frac{2x\sin^2\frac{t}{2}}{1-2x\cos t+x^2}\right]_0^\pi + \int_0^\pi \chi(t)\frac{d}{dt}\left(\frac{2x\sin^2\frac{t}{2}}{1-2x\cos t+x^2}\right)dt.$$

Now, since as $t\to 0$, we have

$$\chi(t) = O\left(\cot\frac{t}{2}\right) = O(t^{-1}),$$

and also $\chi(\pi)=0$, we get

$$F(x) = \int_0^\pi \chi(t)\frac{x(1-x)^2\sin t}{(1-2x\cos t+x^2)^2}\,dt.$$

Suppose the integral (1.1) diverges to $+\infty$. Then corresponding to an arbitrarily great positive number M, an interval $(0,\eta)$ of t can be found in which $\chi(t)>M$. Let A be the maximum absolute value of $\chi(t)$ in (η,π). Then

$$F(x) > M\int_0^\eta \frac{x(1-x)^2\sin t}{(1-2x\cos t+x^2)^2}\,dt - A\int_\eta^\pi \frac{x(1-x)^2\sin t}{(1-2x\cos t+x^2)^2}\,dt$$

$$= M\frac{2x\sin^2\frac{\eta}{2}}{1-2x\cos\eta+x^2} - A\left\{\frac{2x}{(1+x)^2} - \frac{2x\sin^2\frac{\eta}{2}}{1-2x\cos\eta+x^2}\right\}$$

$$= (M+A)\frac{2x\sin^2\frac{\eta}{2}}{1-2x\cos\eta+x^2} - A\frac{2x}{(1+x)^2}.$$

Hence we have

$$\lim_{x\to 1} F(x) \geqq \frac{1}{2}M,$$

which proves the theorem.

The case in which the integral (1.1) diverges to $-\infty$, can be treated in a similar manner.

References.

1. G. H. Hardy *and* J. E. Littlewood, "The allied series of a Fourier series", *Proc. London Math. Soc.* (2), 24 (1925), pp. 211–246.
2. R. E. A. C. Paley, "On the Cesàro summability of Fourier series and allied series", *Proc. Camb. Phil. Soc.*, 26 (1930), pp. 173–203.
3. B. N. Prasad, "A theorem for the Cesàro summability of the allied series of a Fourier series", *Journal London Math. Soc.*, 6 (1931), pp. 274–278.

ON A METHOD OF SUMMATION OF FOURIER SERIES.*

By A. F. Moursund.

1. **Introduction.** In a paper published in 1921 F. Nevanlinna sets up and discusses an interesting method of summation.[1] In this paper we consider the effectiveness of his method (the N_q-method) for the summation of Fourier series.

2. **The N_q-method.** Let $F(n)$ be a function of n defined for all values of $n>0$ and let it be desired to consider a "generalized limit" of $F(n)$ as $n\to\infty$. The N_q-summation method consists in forming the N_q-transform or mean

$$N_q F(n) \equiv \int_0^1 q(t)\, F(nt)\, dt$$

and then considering the limit

$$\lim_{n\to\infty} N_q F(n),$$

where

(1) $q(t)$ is non-negative and monotone increasing for $0<t<1$;

(2) $\int_0^1 q(t)\, dt = 1$;

(3) $q(t)$ is such that the limit

$$\lim_{\delta\to 0} \int_0^{1-\delta} q(t) \log \frac{1}{1-t}\, dt$$

exists.

When the limit of $N_q F(n)$ as $n\to\infty$ exists it will be called the N_q-limit of $F(n)$. It is essential for the application of the N_q-method that $F(n)$ be defined for almost all values of $n>0$, hence if this function is defined only for discrete values of n, for example the nth partial sum of a series, it is necessary to define it for the intermediary values by some method such as linear interpolation.

For the convenience of the reader we state Nevanlinna's results in the form of lemmas.

Lemma 2.1. *The N_q-method is regular and preserves uniformity of approach to a limit.*

Lemma 2.2. *If the kernel*

$$q(t) \equiv \alpha(1-t)^{\alpha-1}, \qquad (0<\alpha\leqq 1),$$

* Received November 18, 1931.—This paper is a part of a dissertation presented for the degree of Doctor of Philosophy at Brown University.

[1] F. Nevanlinna, Über die Summation der Fourier'schen Reihen und Integrale, Översikt av Finska Vetenskaps-Societetens Förhandlingar, 64, A, No. 3, 14 pp. (1921–1922).

then the N_q-method reduces to the Riesz method, (R, α), and is consequently essentially equivalent to the Cesàro method, (C, α).

LEMMA 2.3. *The N_q-method is effective for summing the Fourier series of any Lebesgue integrable function $f(x)$ to $f(x)$ at all points x for which*

$$f(x+s)+f(x-s)-2f(x) = o(1);$$

moreover the approach to $f(x)$ is uniform in the interior of any interval of continuity of $f(x)$.

We add the following lemma which will be useful later.

LEMMA 2.4. *If $F(n)$ is such that the double integral*

$$\int_0^1 q_2(t)\,dt \int_0^1 q_1(s)\,|F(nts)|\,ds$$

exists, then the two N_q-limiting processes, N_{q_1} and N_{q_2}, are permutable, i. e.

$$N_{q_2} N_{q_1} F(n) = N_{q_1} N_{q_2} F(n).$$

Proof. By the well known Fubini theorem we have

$$\int_0^1 q_2(t)\,dt \int_0^1 q_1(s)\,F(nts)\,ds = \int_0^1 q_1(s)\,ds \int_0^1 q_2(t)\,F(nts)\,dt.$$

3. Notation. The following notation will be used throughout the paper.

We consider only functions $f(x)$ which are periodic of period 2π and either Lebesgue integrable $[f(x) \subset L]$ or of bounded variation $[f(x) \subset B]$. We set

$$a_n \equiv \frac{1}{\pi}\int_{-\pi}^{\pi} f(s)\cos n s\,ds,$$

$$b_n \equiv \frac{1}{\pi}\int_{-\pi}^{\pi} f(s)\sin n s\,ds.$$

Then $f(x)$ generates the Fourier series

$$\frac{a_0}{2}+\sum_{n=1}^{\infty}(a_n\cos n x+b_n\sin n x)$$

whose conjugate series is

$$\sum_{n=1}^{\infty}(-b_n\cos n x+a_n\sin n x).$$

We designate by $s_n(x)$, $\sigma_n(x)$, $s_n'(x)$, and $\sigma_n'(x)$ the sum of the first n terms of the Fourier series, conjugate Fourier series, derived series of the Fourier series, and derived series of the conjugate Fourier series, respectively; and by $S_n(x)$ the nth Fejér partial sum of the Fourier series. The same symbols preceded by N_q are then the corresponding Nevanlinna transforms.

We set

$$\varphi(s) \equiv f(x+s)+f(x-s)-2f(x),$$

$$\Phi(s) \equiv \int_0^s |\varphi(t)|\,dt,$$

(3.1)

$$\Phi^*(s) \equiv \int_0^s \varphi(t)\,dt,$$

$$F(s) \equiv \int_0^s |\,d\varphi(t)\,|;$$

$$\psi(s) \equiv f(x+s)-f(x-s)-2sf'(x),$$

(3.2)

$$G(s) \equiv \int_0^s |\,d\,\psi(t)\,|;$$

$$\psi^*(s) \equiv f(x+s)-f(x-s),$$

(3.3)

$$\Psi^*(s) \equiv \int_0^s |\,\psi^*(t)\,|\,dt.$$

We say that with respect to $f(x)$ a point x is:

Def. 3.1. (F) *regular if* $\varphi(s) = o(1)$, *as* $s \to 0$.

Def. 3.2. (L) *regular if* $\Phi(s) = o(s)$, *as* $s \to 0$.

Def. 3.3. $(\tilde{F})$ *regular if* (i) $f(x)$ *is continuous and* (ii)

$$\tilde{f}(x) \equiv \frac{-1}{2\pi} \lim_{\varepsilon \to 0} \int_\varepsilon^\pi \psi^*(s) \cot \frac{s}{2}\,ds$$

exists and is finite.

Def. 3.4. $(\tilde{L})$ *regular if* (i) $\Psi^*(s) = o(s)$, *as* $s \to 0$, *and* (ii) $\tilde{f}(x)$ *exists and is finite.*

Def. 3.5. (L') *regular if* (i) $f'(x)$ *exists and* (ii) $G(s) = o(s)$, *as* $s \to 0$.

Def. 3.6. $(\tilde{L}')$ *regular if* (i) *the limit*

$$\tilde{f}'(x) \equiv -\frac{1}{4\pi} \lim_{\varepsilon \to 0} \int_\varepsilon^\pi \varphi(s) \left[\sin \frac{s}{2}\right]^{-2} ds$$

exists and is finite and (ii) $F(s) = o(s)$, *as* $s \to 0$.

It is well known[2] that, except for sets of measure zero, all points are (L) and $(\tilde{L})$ regular for $f(x) \subset L$ and (L') and $(\tilde{L}')$ regular for $f(x) \subset B$. In particular all (F) regular points are (L) regular and all $(\tilde{F})$ regular points are $(\tilde{L})$ regular.

We say that a summation method is:

Def. 3.7 [3.8]. (F) *effective* [(L) *effective*] *if it sums the Fourier series of* $f(x) \subset L$ *to* $f(x)$ *at the* (F) *regular* [(L) *regular*] *points.*

Def. 3.9 [3.10]. $(\tilde{F})$ *effective* [$(\tilde{L})$ *effective*] *if it sums the conjugate Fourier series of* $f(x) \subset L$ *to* $\tilde{f}(x)$ *at the* $(\tilde{F})$ *regular* [$(\tilde{L})$ *regular*] *points.*

[2] Hobson: The Theory of Functions of a Real Variable, vol. II. A. Plessner, Zur Theorie der konjugierten trigonometrischen Reihen, Mitteilungen deš Mathematischen Seminars der Universität Giessen, Heft 10 (1923).

Def. 3.11. *(L') effective if it sums the derived series of the Fourier series of $f(x) \subset B$ to $f'(x)$ at the (L') regular points.*

Def. 3.12. *$(\tilde{L}')$ effective if it sums the derived series of the conjugate Fourier series of $f(x) \subset B$ to $\tilde{f}'(x)$ at the $(\tilde{L}')$ regular points.*

4. The Nevanlinna transforms. We have

$$s_n(x) = \frac{a_0}{2} + \sum_1^n (a_n \cos nx + b_n \sin nx)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \frac{\sin\left(n+\frac{1}{2}\right)(x-s)}{\sin\frac{x-s}{2}}\, ds$$

$$= \frac{1}{2\pi} \int_0^{\pi} [f(x+s)+f(x-s)] \frac{\sin\left(n+\frac{1}{2}\right)s}{\sin\frac{s}{2}}\, ds \tag{4.1}$$

$$= f(x) + \frac{1}{2\pi} \int_0^{\pi} \varphi(s) \frac{\sin\left(n+\frac{1}{2}\right)s}{\sin\frac{s}{2}}\, ds.$$

Hence

$$N_q s_n(x) = f(x) + \frac{1}{2\pi} \int_0^1 q(t)\, dt \int_0^{\pi} \varphi(s) \frac{\sin\left(nt+\frac{1}{2}\right)s}{\sin\frac{s}{2}}\, ds$$

$$= f(x) + \frac{1}{2\pi}\left[\int_0^1 \int_0^{\pi/n} + \int_0^1 \int_{\pi/n}^{\delta} + \int_0^1 q(t)\, dt \int_{\delta}^{\pi} \varphi(s) \sin nts \cot\frac{s}{2}\, ds\right] \tag{4.2}$$

$$+ \frac{1}{2\pi} \int_0^1 q(t)\, dt \int_0^{\pi} \varphi(s) \cos nts\, ds$$

$$\equiv f(x) + K_1 + K_2 + K_3 + K_4.$$

Also

$$\sigma_n(x) = \sum_1^n (-b_n \cos nx + a_n \sin nx)$$

$$= \frac{1}{\pi} \sum_1^n \int_{-\pi}^{\pi} f(t) \sin n(x-t)\, dt \tag{4.3}$$

$$= \frac{1}{\pi} \sum_1^n \int_0^{\pi} [f(x-s)-f(x+s)] \sin ns\, ds$$

$$= -\frac{1}{\pi}\int_0^\pi \psi^*(s)\frac{\cos\frac{s}{2}-\cos\left(n+\frac{1}{2}\right)s}{2\sin\frac{s}{2}}\,ds$$

$$= -\frac{1}{2\pi}\int_0^\pi \psi^*(s)\cot\frac{s}{2}\,[1-\cos ns]\,ds$$

$$-\frac{1}{2\pi}\int_0^\pi \psi^*(s)\sin ns\,ds,$$

whence

$$\begin{aligned} N_q\,\sigma_n(x) &= -\frac{1}{2\pi}\int_0^1 q(t)\,dt\int_0^\pi \psi^*(s)\cot\frac{s}{2}\,[1-\cos nst]\,ds \\ &\quad -\frac{1}{2\pi}\int_0^1 q(t)\,dt\int_0^\pi \psi^*(s)\sin nst\,dt \\ &= -\frac{1}{2\pi}\int_{\pi/n}^\pi \psi^*(s)\cot\frac{s}{2}\,ds \\ &\quad -\frac{1}{2\pi}\int_0^1 q(t)\,dt\int_0^{\pi/n} \psi^*(s)\cot\frac{s}{2}\,[1-\cos nst]\,ds \\ &\quad +\frac{1}{2\pi}\left[\int_0^1\int_{\pi/n}^{\delta} + \int_0^1 q(t)\,dt\int_{\delta}^{\pi}\psi^*(s)\cot\frac{s}{2}\cos nst\,ds\right] \\ &\quad -\frac{1}{2\pi}\int_0^1 q(t)\,dt\int_0^\pi \psi^*(s)\sin nst\,dt \\ &\equiv -\frac{1}{2\pi}\int_{\pi/n}^\pi \psi^*(s)\cot\frac{s}{2}\,ds + L_1 + L_2 + L_3 + L_4. \end{aligned} \tag{4.4}$$

It is readily seen that

$$\begin{aligned} s_n'(x) &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(s)\frac{\partial}{\partial x}\left[\frac{\sin\left(n+\frac{1}{2}\right)(x-s)}{2\sin\frac{x-s}{2}}\right]ds \\ &= -\frac{1}{\pi}\int_{-\pi}^{\pi} f(s)\frac{\partial}{\partial s}\left[\frac{\sin\left(n+\frac{1}{2}\right)(x-s)}{2\sin\frac{x-s}{2}}\right]ds \\ &= -\frac{1}{\pi}\int_0^\pi [f(x+s)-f(x-s)]\,\frac{\partial}{\partial s}\left[\frac{\sin\left(n+\frac{1}{2}\right)s}{2\sin\frac{s}{2}}\right]ds \\ &= f'(x)+\frac{1}{2\pi}\int_0^\pi \frac{\sin\left(n+\frac{1}{2}\right)s}{\sin\frac{s}{2}}\,d\psi(s) \end{aligned} \tag{4.5}$$

wherever $f'(x)$ exists, whence

$$N_q s_n'(x) = f'(x) + \frac{1}{2\pi}\int_0^1 q(t)\,dt \int_0^\pi \frac{\sin\left(nt+\frac{1}{2}\right)s}{\sin\frac{s}{2}}\,d\psi(s)$$

$$= f'(x) + \frac{1}{2\pi}\left[\int_0^1\int_0^{\pi/n} + \int_0^1\int_{\pi/n}^{\delta} + \int_0^1\int_{\delta}^{\pi}\right] \tag{4.6}$$

$$\equiv f'(x) + M_1 + M_2 + M_3.$$

In an analogous manner,

$$\sigma_n'(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(s)\frac{\partial}{\partial x}\left[\frac{\cos\frac{x-s}{2} - \cos\left(n+\frac{1}{2}\right)(x-s)}{\sin\frac{x-s}{2}}\right]ds$$

$$= -\frac{1}{2\pi}\int_{-\pi}^{\pi} f(s)\frac{\partial}{\partial s}\left[\frac{\cos\frac{x-s}{2} - \cos\left(n+\frac{1}{2}\right)(x-s)}{\sin\frac{x-s}{2}}\right]ds$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\cos\frac{x-s}{2} - \cos\left(n+\frac{1}{2}\right)(x-s)}{\sin\frac{x-s}{2}}\,df(s) \tag{4.7}$$

$$= -\frac{1}{2\pi}\int_0^{\pi} \frac{\cos\frac{s}{2} - \cos\left(n+\frac{1}{2}\right)s}{\sin\frac{s}{2}}\,d\varphi(s),$$

and

$$N_q\,\sigma_n'(x) = -\frac{1}{2\pi}\int_0^1 q(t)\,dt\int_0^\pi \frac{\cos\frac{s}{2} - \cos\left(nt+\frac{1}{2}\right)s}{\sin\frac{s}{2}}\,d\varphi(s)$$

$$= -\frac{1}{2\pi}\int_{\pi/n}^{\pi} \cot\frac{s}{2}\,d\varphi(s)$$

$$-\frac{1}{2\pi}\int_0^1 q(t)\,dt\int_0^{\pi/n} \frac{\cos\frac{s}{2} - \cos\left(nt+\frac{1}{2}\right)s}{\sin\frac{s}{2}}\,d\varphi(s) \tag{4.8}$$

$$+\frac{1}{2\pi}\left[\int_0^1\int_{\pi/n}^{\delta} + \int_0^1 q(t)\,dt\int_{\delta}^{\pi} \frac{\cos\left(nt+\frac{1}{2}\right)s}{\sin\frac{s}{2}}\,d\varphi(s)\right]$$

$$\equiv -\frac{1}{2\pi}\int_{\pi/n}^{\pi}\cot\frac{s}{2}\,d\varphi(s) + N_1 + N_2 + N_3.$$

Finally,[3]

$$
\begin{aligned}
S_n(x) &= f(x)+\frac{1}{n\pi}\int_0^{\pi/2}\varphi(s)\left(\frac{\sin ns}{\sin s}\right)^2 ds \\
&= f(x)+\frac{1}{n\pi}\,\Phi^*\left(\frac{\pi}{2}\right)\sin^2\frac{n\pi}{2} \\
&\quad +\frac{2}{n\pi}\int_0^{\pi/2}\Phi^*(s)\cot s\left(\frac{\sin ns}{\sin s}\right)^2 ds \\
&\quad -\frac{1}{\pi}\int_0^{\pi/2}\frac{\Phi^*(s)}{\sin s}\cdot\frac{\sin 2ns}{\sin s}\,ds \\
&\equiv f(x)+P_1(x,n)+P_2(x,n)+P_3(x,n),
\end{aligned}
\tag{4.9}
$$

and

$$
\begin{aligned}
N_q S_n(x) &= f(x)+\int_0^1 q(t)\,P_1(x,nt)\,dt \\
&\quad +\int_0^1 q(t)\,P_2(x,nt)\,dt+\int_0^1 q(t)\,P_3(x,nt)\,dt \\
&\equiv f(x)+R_1+R_2+R_3.
\end{aligned}
\tag{4.10}
$$

5. **Some preliminary lemmas.** We assume, as everywhere in this paper, that the kernel $q(t)$ of the method N_q satisfies the three conditions given in section 2. We set

$$Q(h)=\int_{1-h}^1 q(t)\,dt, \qquad h=\frac{\pi}{ns}\leqq 1.$$

We have then:

LEMMA 5.1. $\left|\int_0^1 \left\{\begin{matrix}\sin\\ \cos\end{matrix}\right\} nst\,q(t)\,dt\right| \leqq Q(h).$

LEMMA 5.2. *$Q(h)$ is a continuous monotone increasing function of h, $Q(0)=0$ and $Q(h)\leqq 1$.*

LEMMA 5.3. *$\int_0^1 \frac{Q(s)}{s}\,ds$ exists.*

Proof of Lemma 5.1. We can write

$$
\begin{aligned}
&\int_0^1 \sin\, nst\,q(t)\,dt \\
&\quad=\int_{-h}^{1-h}\sin\, ns(t+h)\,q(t+h)\,dt \\
&\quad=-\int_{-h}^{1-h}\sin\, nst\,q(t+h)\,dt \\
&\quad=\frac{1}{2}\int_0^1 \sin\, nst\,q(t)\,dt-\frac{1}{2}\int_{-h}^{1-h}\sin\, nst\,q(t+h)\,dt
\end{aligned}
$$

[3] Hobson, loc. cit., p. 557, pp. 563–564.

$$= \frac{1}{2}\int_{1-h}^{1} \sin\, nst\, q(t)\, dt + \frac{1}{2}\int_{0}^{1-h} \sin\, nst\, [q(t)-q(t+h)]\, dt$$
$$-\frac{1}{2}\int_{-h}^{0} \sin\, nst\, q(t+h)\, dt.$$

Then

$$\left|\int_0^1 \sin\, nst\, q(t)\, dt\right|$$
$$\leqq \frac{1}{2}\int_{1-h}^{1} q(t)\, dt + \frac{1}{2}\int_{0}^{1-h} [q(t+h)-q(t)]\, dt + \frac{1}{2}\int_{-h}^{0} q(t+h)\, dt = Q(h).$$

Throughout this proof sin may be replaced by cos.

Lemma 5.2 follows from the definition of $Q(h)$.

Lemma 5.3 follows from condition (3) on $q(t)$ [section 2], for we have

$$\int_0^1 \frac{Q(s)}{s}\, ds = \int_0^1 \frac{1}{s}\, ds \int_{1-s}^{1} q(t)\, dt$$
$$= \int_0^1 q(t)\, dt \int_{1-t}^{1} \frac{1}{s}\, ds = \int_0^1 q(t)\, \log \frac{1}{1-t}\, dt.$$

6. **The effectiveness of the N_q-method.** In this section we will state our principal theorems and give their proofs.

THEOREM 6.1. *The N_q-method is (F) and (L) effective.*

Proof. At the (L) regular points we have, from (4.2):

$$|K_1| \leqq \frac{1}{2\pi}\int_0^1 q(t)\, dt \int_0^{\pi/n} |\varphi(s)|\, |\sin\, nts|\, \frac{2}{s}\, ds \leqq \frac{n}{\pi}\, \Phi\left(\frac{\pi}{n}\right) = o(1);$$

also by an interchange of the order of integration, integration by parts, and Lemmas 5.1, 5.2, and 5.3,

$$|K_2| = \frac{1}{2\pi}\left|\int_{\pi/n}^{\delta} \varphi(s)\, \cot\frac{s}{2}\, ds \int_0^1 q(t)\, \sin\, nts\, dt\right|$$
$$\leqq \frac{1}{\pi}\int_{\pi/n}^{\delta} |\varphi(s)|\, \frac{1}{s}\, Q\left(\frac{\pi}{ns}\right) ds$$
$$= \frac{1}{\pi}\, Q\left(\frac{\pi}{ns}\right)\frac{1}{s}\, \Phi(s)\Big]_{\pi/n}^{\delta} - \frac{1}{\pi}\int_{\pi/n}^{\delta} \frac{1}{s}\, \Phi(s)\, Q'\left(\frac{\pi}{ns}\right) ds$$
$$+ \frac{1}{\pi}\int_{\pi/n}^{\delta} \frac{1}{s}\, \Phi(s)\, \frac{Q\left(\frac{\pi}{ns}\right)}{s}\, ds$$
$$\leqq o(1)+o(1)+o(1)\int_0^1 \frac{Q(s)}{s}\, ds = o(1), \text{ as } \delta \to 0;$$

and finally by the Riemann-Lebesgue theorem and Lemma 2.1,

$$|K_3| = o(1), \text{ for a fixed } \delta,$$
$$|K_4| = o(1).$$

Hence

$$N_q s_n(x) \to f(x) \text{ as } n \to \infty.$$

The (F) effectiveness follows from the (L) effectiveness since the (F) regular points are a subset of the (L) regular points. Uniformity of approach in any interval of continuity (see Lemma 2.3) follows from Lemma 2.1 and the fact that in such an interval $\Phi(s) = o(s)$ uniformly with respect to x.

THEOREM 6.2. *The N_q-method is $(\tilde{F})$ and $(\tilde{L})$ effective.*

Proof. At the $(\tilde{L})$ regular points we have, from (4.4):

$$|L_1| \leqq \frac{1}{2\pi} \int_0^1 q(t)\,dt \int_0^{\pi/n} |\psi^*(s)| \frac{2}{s} \cdot 2 \sin^2 \frac{nst}{2}\,ds \leqq \frac{n}{\pi} \psi\left(\frac{\pi}{n}\right) = o(1);$$

and by computations like those in the proof of Theorem 6.1 for the K's having the same subscripts,

$$|L_2| = o(1), \text{ as } \delta \to 0,$$
$$|L_3| = o(1), \text{ for a fixed } \delta,$$

and

$$|L_4| = o(1).$$

Hence

$$N_q\, \sigma_n(x) = -\frac{1}{2\pi} \int_{\pi/n}^{\delta} \psi^*(s) \cot \frac{s}{2}\,ds + o(1) \to \tilde{f}(x) \text{ as } n \to \infty.$$

The $(\tilde{F})$ effectiveness follows from the fact that the $(\tilde{F})$ regular points are a subset of the $(\tilde{L})$ regular points.

THEOREM 6.3. *The N_q-method is (L') effective.*

Proof. At the (L') regular points we have from (4.6):

$$|M_1| \leqq \frac{1}{2\pi} \int_0^1 q(t)\,dt \int_0^{\pi/n} \left| \frac{\sin\left(nt + \frac{1}{2}\right)s}{\sin \frac{s}{2}} \right| |d\psi(s)|$$

$$\leqq \frac{2n+1}{\pi} \int_0^1 q(t)\,dt\, G\left(\frac{\pi}{n}\right) = o(1);$$

$$|M_2| \leqq \frac{1}{2\pi} \left| \int_{\pi/n}^{\delta} \cot \frac{s}{2}\,d\psi(s) \int_0^1 q(t) \sin nts\,dt \right|$$

$$+ \frac{1}{2\pi} \left| \int_{\pi/n}^{\delta} d\psi(s) \int_0^1 q(t) \cos nts\,dt \right|$$

$$\leqq \frac{1}{2\pi}\int_{\pi/n}^{\delta}\left(\frac{2}{s}+1\right)Q\left(\frac{\pi}{ns}\right)|d\psi(s)|$$

$$\leqq \frac{2+\delta}{2\pi}\int_{\pi/n}^{\delta}Q\left(\frac{\pi}{ns}\right)\frac{1}{s}|d\psi(s)|$$

$$\leqq \frac{2+\delta}{2\pi}\left\{Q\left(\frac{\pi}{ns}\right)\frac{1}{s}G(s)\Big]_{\pi/n}^{\delta}-\int_{\pi/n}^{\delta}\frac{1}{s}G(s)\,Q'\left(\frac{\pi}{ns}\right)ds\right.$$

$$\left.+\int_{\pi/n}^{\delta}\frac{1}{s}G(s)Q\left(\frac{\pi}{ns}\right)ds\right\}=o(1),\ \text{as}\ \delta\to 0,$$

as in the proof of $K_2 = o(1)$; and finally

$$|M_3| \leqq \frac{1}{2\pi}\int_{\delta}^{\pi}\left(\cot\frac{s}{2}+1\right)Q\left(\frac{\pi}{ns}\right)|d\psi(s)|$$

$$\leqq \left(\cot\frac{\delta}{2}+1\right)\frac{1}{2\pi}Q\left(\frac{\pi}{n\delta}\right)G(\pi)=o(1),\ \text{for a fixed}\ \delta.$$

Hence

$$N_q\, s_n'(x)\to f'(x)\ \text{as}\ n\to\infty.$$

THEOREM 6.4. *The N_q-method is $(\tilde{L}')$ effective.*

Proof. At the $(\tilde{L}')$ regular points we have from (4.8):

$$|N_1| \leqq \frac{1}{2\pi}\int_0^1 q(t)\,dt\int_0^{\pi/n}(2n+1)|d\varphi(s)| = \frac{2n+1}{2\pi}F\left(\frac{\pi}{n}\right)=o(1),$$

since

$$\left|\frac{\cos\frac{s}{2}-\cos\left(nt+\frac{1}{2}\right)s}{\sin\frac{s}{2}}\right| \leqq \left|\frac{\cos\frac{s}{2}(1-\cos nts)}{\sin\frac{s}{2}}\right|+|\sin nts|$$

$$\leqq \left|\frac{2}{s}\cdot 2\sin^2\frac{nts}{2}\right|+1 \leqq 2n+1;$$

and

$$|N_2| = o(1),\ \text{as}\ \delta\to 0,$$
$$|N_3| = o(1),\ \text{for a fixed}\ \delta,$$

as in the proofs of $|M_2| = o(1)$ and $|M_3| = o(1)$.

Integrating by parts we have,

$$-\frac{1}{2\pi}\int_{\pi/n}^{\pi}\cot\frac{s}{2}\,d\varphi(s) = -\frac{1}{2\pi}\cot\frac{s}{2}\,\varphi(s)\Big]_{\pi/n}^{\pi}$$

$$-\frac{1}{4\pi}\int_{\pi/n}^{\pi}\frac{\varphi(s)}{\sin^2\frac{s}{2}}\,ds\to\tilde{f}'(x)\ \text{as}\ n\to\infty,$$

since at the $(\tilde{L}')$ regular points the term containing the integral tends to $\tilde{f}'(x)$, a finite limit for the existence of which it is necessary that

$$\frac{\varphi\left(\frac{\pi}{n}\right)}{\sin\frac{\pi}{2n}} = o(1).$$

Hence

$$N_q\,\sigma_n'(x) \to \tilde{f}'(x) \text{ as } n \to \infty.$$

7. THEOREM 7.1. *If the function $f(x) \subset L$ and is periodic of period 2π, then the $(C, 1)\cdot N_q$ sum of its Fourier series is equal to $f(x)$ at each point x for which $\Phi^*(s) = o(s)$, as $s \to 0$.*

Proof. At a point x for which

$$\Phi^*(s) = o(s)$$

we have

$$\frac{\Phi^*(s)}{\sin s} = o(1),$$

and then from (4.9) and (4.10) by Lemma 2.1:

$$R_1 = o(1),$$

since obviously $P_1(x, n) = o(1)$; also

$$R_2 = o(1),$$

since $P_2(x, n) = o(1)$ being the Fejér partial sum of the continuous function $\cot\frac{X}{2}\,\Phi^*\left(\frac{X}{2}\right)$ at the point $X = 0$. Finally

$$R_3 = \int_0^1 \int_0^{\pi/2n} + \int_0^1 \int_{\pi/2n}^{\delta} + \int_0^1 \int_{\delta}^{\pi/2} \equiv T_1 + T_2 + T_3,$$

and we have immediately

$$\begin{aligned} |T_1| &\leqq \frac{1}{\pi}\int_0^1 q(t)\,dt \int_0^{\pi/2n} \frac{|\Phi^*(s)|}{\sin s}\,\frac{|\sin 2nts|}{\sin s}\,ds \\ &\leqq \frac{1}{\pi}\int_0^1 q(t)\,dt \cdot 4n \int_0^{\pi/2n} \frac{|\Phi^*(s)|}{\sin s}\,ds \\ &= \frac{4n}{\pi}\cdot o(1)\cdot\frac{\pi}{2n} = o(1); \end{aligned}$$

also by an interchange of the order of integration and Lemmas 5.1, 5.2, and 5.3

$$|T_2| = \frac{1}{\pi}\left|\int_{\pi/2n}^{\delta} \frac{\Phi^*(s)}{\sin s}\cdot\frac{1}{\sin s}\,ds\int_0^1 q(t)\sin 2nts\,dt\right|$$

$$\leqq \frac{1}{\pi}\int_{\pi/2n}^{\delta}\frac{\Phi^*(s)}{\sin s}\cdot\frac{-s}{\sin s}\cdot\frac{1}{s}\,Q\left(\frac{\pi}{2ns}\right)ds$$

$$\leqq \frac{2}{\pi}\cdot o(1)\cdot\int_{\pi/2n}^{\delta}\frac{Q\left(\frac{\pi}{2ns}\right)}{s}\,ds$$

$$\leqq \left[\frac{2}{\pi}\int_0^1\frac{Q(s)}{s}\,ds\right]\cdot o(1) = o(1), \text{ as } \delta\to 0,$$

and

$$|T_3| = \frac{1}{\pi}\left|\int_{\delta}^{\pi/2}\frac{\Phi^*(s)}{\sin^2 s}\,ds\int_0^1 q(t)\sin 2nts\,dt\right|$$

$$\leqq \frac{1}{\pi}\int_{\delta}^{\pi/2}\frac{|\Phi^*(s)|}{\sin^2 s}\,Q\left(\frac{\pi}{2ns}\right)ds$$

$$= o(1)\cdot\frac{1}{\pi}\int_{\delta}^{\pi/2}\frac{\Phi^*(s)}{\sin^2 s}\,ds = o(1), \text{ for a fixed } \delta.$$

Hence

$$N_q S_n \to f(x) \text{ as } n\to\infty.$$

By Lemma 2.4 the order of application of the two transformations is immaterial.

We may choose $q(t) = \varepsilon(1-t)^{\varepsilon-1}$ and obtain as a special case of our theorem the usual theorem on summability $(C, 1+\varepsilon)$.[4]

[4] We plan to consider the summation of the pth derived series of the Fourier series in a second paper.

REMARKS ON A PROBLEM OF TOEPLITZ.[1]

By F. Bohnenblust and E. Hille.

In this note we desire to make some additions and corrections to our paper "On the absolute convergence of Dirichlet series".[2] The additions refer to the following problem of Toeplitz:[3]

Let $\mathrm{Max}\,|Q(x)|$ denote the maximum of the absolute value of the quadratic form

$$Q(x) \equiv \sum_{i,k=1}^{n} a_{ik}\, x_i\, x_k$$

in the unit cube $|x_i| \leqq 1$. We normalize these quadratic forms by making the coefficients symmetric $a_{ik} = a_{ki}$, and the sum of their absolute values equal to one: $\sum |a_{ik}| = 1$. Considering all normalized quadratic forms we determine the one, $Q_0(x)$, whose maximum is the least possible and put

$$\varphi(n) = \frac{1}{\mathrm{Max}\,|Q_0(x)|} = \mathrm{Max}\left(\frac{1}{\mathrm{Max}\,|Q(x)|}\right);$$

thus defining a number theoretic function $\varphi(n)$. *Toeplitz's problem is the study of this function.* He showed $\varphi(2) = \sqrt{2}$, and $\varphi(n) \leqq n$, while $\varphi(n) \geqq \sqrt{n}$ for certain n tending to infinity.

There exists naturally a similar problem for normalized m-ic forms

$$Q_m(x) \equiv \sum_{i_1 \cdots i_m = 1}^{n} a_{i_1 \cdots i_m}\, x_{i_1} \cdots x_{i_m};$$

we denote the corresponding function of the dimension n by

$$\varphi_m(n) = \mathrm{Max}\left(\frac{1}{\mathrm{Max}\,|Q_m(x)|}\right).$$

We have been unable to find an exact expression for the non-decreasing function $\varphi_m(n)$, but the order of its growth follows immediately from our results. To be precise we prove

$$\varphi_m(n) = O(n^{(m-1)/2}) \text{ and } [\varphi_m(n)]^{-1} = O(n^{-(m-1)/2});$$

---

[1] Received April 19, 1932.

[2] These Annals, (2) 32, (1931), pp. 600–622.

[3] O. Toeplitz, Über eine bei den Dirichletschen Reihen auftretende Aufgabe aus der Theorie der Potenzreihen von unendlich vielen Veränderlichen, Gött. Nachr. (1913), pp. 417–432.

where the constants involved in these O-results depend only on m. The first of these evaluations follows readily from equation (1.12) on page 608 of our paper:

$$\text{Max } |Q_m(x)| \geqq A_m \Big\{ \sum_{i_1 \cdots i_m} |a_{i_1 \cdots i_m}|^{\varrho} \Big\}^{1/\varrho}; \quad \varrho = \frac{2m}{m+1};$$

for the left hand side of this inequality reaches a minimum as the coefficients vary subject to the condition $\sum |a_{i_1 \cdots i_m}| = 1$, when each is equal to n^{-m}. The minimal value is $n^{-(m-1)/2}$ and hence

$$\text{Max } |Q_m(x)| \geqq A_m \, n^{-(m-1)/2},$$

proving the first assertion.

The second result is equivalent to equation (3.3) on page 612 for all n which are powers of the chosen prime $p > m$. But it can be extended to cover all integers n, due to the monotonic character of $\varphi_m(n)$ and the sufficient density of p^k.

The corrections are the following ones:

1) On page 615, lines 7 and 9 read $p^{\vartheta(m-1)/2}$ instead of p^{ϑ}.

2) On page 622. In the remark at the end of the paper, take the function $\zeta(\frac{1}{2} + \sigma + s)$ instead of $\zeta(1 - \sigma + s)$ in order to obtain a function for which $\sigma_a - \sigma_u = \sigma$.

PRINCETON UNIVERSITY.

A NOTE ON TAUBERIAN THEOREMS.[1]

By Norbert Wiener.

The following is a list of corrections to my paper Tauberian Theorems in this volume of the Annals, pp. 1–100.

Page 3, formula (0.07). Read

$$\lim_{x\to 1-0} \sum_0^\infty a_n x^n = A$$

for $\lim_{x\to 1-0} \frac{1}{1-x} \sum_0^\infty a_n x^n = A.$

Page 4, formula (0.12). Read

$$\lim_{x\to 1-0} (1-x) \sum_0^\infty \frac{n a_n x^n}{1-x^n} = A$$

for $\lim_{x\to 1-0} \sum_0^\infty \frac{n a_n x^n}{1-x^n} = A.$

Page 20, line following formula (2.63). Read x for λ.

Page 24, formula (3.225), second line. Read

$$\leqq \sum_{k=-\infty}^{+\infty} \max_z \max_{k\leqq x\leqq k+1} |f(x+z)| \lim_{N\to\infty} \frac{1}{\pi N} \int_{-\infty}^{+\infty} \Phi_2(y)\, dy$$

for $\leqq \sum_{k=-\infty}^{+\infty} \max_{k\leqq x\leqq k+1} |f(x+y)| \lim_{N\to\infty} \frac{1}{\pi N} \int_{-\infty}^{+\infty} \Phi_2(y)\, dy.$

Page 28, line 9. Delete "not be equivalent to 0, and let it".

Page 40, formula (9.06). Read $A(n)$ for A_n.

Page 45, formula (10.15). In the second line, delete the minus sign before the integral.

Page 47, formula (10.26). Insert a — after the sign of equality.

Page 47, line 10. For $2(\cos Bx - 1)$ read $2(\cos Bx - 1)/Bx^2$.

Page 66, formula (18.15). Insert du after $M(u)$.

Page 67, formula (18.22). Insert du after $M^*(u)$.

Page 72, formula (19.35). Read $\varphi(z) = \frac{e^{z^2/8}}{z\sqrt{2}}$.

Page 81, line 4. Delete "the same", and insert in its place "implied by this boundedness".

Page 84. The phrase "and hence by an application of (22.32)" should be raised from the line following (22.321) to a position directly to the left of the last line of that formula.

Page 86, line 2. The expression on this line should be squared.

Page 86, formula (22.48). The expression in the first line of this formula should be squared.

Page 88, formula (22.60). The formula should end "$= 0$".

[1] Received April 18, 1932.

NOTE ON A PRECEDING PAPER.[1]

By George D. Birkhoff.

In a preceding number of these *Annals*[2] in an article entitled "A Set of Postulates for Plane Geometry Based on Scale and Protractor", the proof of the second part of Theorem 5, correctly given for right triangles, was incorrectly extended to the general triangle.

Dr. Dorroh had previously called my attention to an alternative proof, but unfortunately I looked over my proof once more too hastily, so that the necessity of revision escaped me. The following short proof is designed to replace my own earlier reasoning.

In any non-degenerate ΔABC let the point B, for instance, tend toward the point C along the line BC. The angle γ at C will remain constant; the angle α at A will vary continuously according to Postulate III, part 2, and will approach 0. During this process the angle α can never become equal to 0 or π, since ΔABC is never degenerate.

Similarly, in view of the first part of Theorem 5, the angle sum is congruent to π in all these positions. It follows, therefore, that the angle β must vary continuously $(\mathrm{mod}\, 2\pi)$ and will tend towards a limiting value β_0. Furthermore, during this variation β can never be equal to 0 or π for the same reasons as before.

But when α is very small and β is very near to β_0, it is clear that $\beta+\gamma$ will differ very little from $\pi(\mathrm{mod}\, 2\pi)$, according to the first part of Theorem 5. This shows that when α is small, the least residues of β and γ must necessarily both lie between 0 and π or between 0 and $-\pi(\mathrm{mod}\, 2\pi)$. Since β and γ, which were any two angles of the triangle, thus have least residues of the same sign, the third part of the theorem is obviously true.

[1] Received May 16, 1932.

[2] This volume, pp. 329–345.

ZUSÄTZE ZUR ARBEIT „ZUR OPERATORENMETHODE ...“[1][2].

Von J. v. Neumann, Princeton.

1. Als erstes soll ein Fehler berichtigt werden, der dem Verf. beim Abfassen der zitierten Arbeit unterlaufen ist. Im Hauptsatz von Teil II, § 3, d. i. Satz 2 (der die Zerlegbarkeit aller Strömungen in ergodische Strömungen aussagt), wird u. a. behauptet, daß die dort konstruierten Mengen $X^{(0)}$ und $M(x)$ m-Räume sind[3], also u. a. vollständig[3]. Dabei wurde übersehen, daß die zu Satz 2 führende Konstruktion die Vollständigkeit nicht sicherstellt. Die Vollständigkeit ist aber insofern wichtig, als z. B. der zweite Teil von Satz 5[4] auf Satz 1 der vorhergehenden Abhandlung des Verf. beruht[5], der seinerseits die Vollständigkeit als Prämisse hat.

Da jeder absolute G_δ-Raum bekanntlich so ummetrisiert werden kann, daß seine Topologie ungeändert bleibt und er vollständig wird[6], würde es genügen, $X^{(0)}$ und die $M(x)$ als G_δ-Mengen zu wählen, um diese Lücke auszufüllen — was in der Tat durchführbar ist. Wir ziehen aber vor, zu zeigen, daß die Vollständigkeit für unsere Zwecke, d. h. für die Gültigkeit von Satz 1, 2 und 3, a. a. O. Anm.[5], entbehrlich ist — d. h. daß der Begriff des „m-Raumes“ ohne sie[7] gefaßt werden kann. Dies ist erwiesen, wenn es uns gelingt, einen metrisch separablen Raum Ω, in dem ein l-Maß μ^* definiert ist, derart zu einem metrischen, separablen und vollständigen Raume $\overline{\Omega}$ mit einem l-Maß $\overline{\mu}^*$ zu erweitern, daß $\overline{\mu}^*$ für Mengen $\subset \Omega$ mit μ^* übereinstimmt und $\overline{\mu}^*(\overline{\Omega}-\Omega)=0$ ist. (Denn dann überträgt sich die Prämisse von Satz 1 ohne weiteres von Ω, Ω' auf $\overline{\Omega}$, $\overline{\Omega}'$ und seine Aussage von $\overline{\Omega}$, $\overline{\Omega}'$ auf Ω, Ω'. Dasselbe gilt für Satz 2 ebendort, allerdings ist dort die Existenz einer eineindeutigen Abbildung von $\overline{\Omega}-\Omega$ auf $\overline{\Omega}'-\Omega'$ notwendig, was z. B. sicher geht, wenn beide die Mächtigkeit des Kontinuums haben[8]. Satz 3 gilt auch, da er direkt aus Satz 2 folgt.)

---

[1] Annals of Math., Bd. 33, 3. Heft (1932), S. 587–642.

[2] Received September 6, 1932.

[3] Vgl. die Definition D_1 auf S. 588 a. a. O.

[4] S. 629 a. a. O.

[5] Annals of Math., Bd. 33, 3. Heft (1932), S. 574–586; vgl. insbesondere Def. 1–5 und Satz 5.

[6] Vgl. z. B. Hausdorff, „Mengenlehre“ (Berlin und Leipzig, 1927), S. 214, Satz III.

[7] D. h. als metrisch und separabel sowie, wie sich w. u. zeigen wird, absolut Borelsch.

[8] Sei C die nirgendsdichte Cantorsche perfekte Menge in ihrer ursprünglichen Topologie, $\overline{\mu}^*$ werde für alle Teilmengen von C gleich 0 definiert. Wir fügen $\overline{\Omega}$, C zu $\overline{\overline{\Omega}}$ zusammen, u. zw. unter Beibehaltung ihrer Topologien und $\overline{\mu}^*$, aber voneinander isoliert. Dann spielt $\overline{\overline{\Omega}}$ die Rolle von $\overline{\Omega}$, und $\overline{\overline{\Omega}}-\Omega \supset C$ hat die Mächtigkeit des Kontinuums.

Wählen wir $\bar{\Omega}$ als vollständige Hülle von Ω[9] und definieren für die $M \subset \bar{\Omega}$ $\bar{\mu}^*(M) = \mu^*(M \cdot \Omega)$ $(M \cdot \Omega \subset \Omega)$, so ist alles Gewünschte erfüllt, bloß bezüglich des l-Maß-Charakters des neuen $\bar{\mu}^*$ sind I., V.[10] zu verifizieren. Dies gelingt in der Tat:

Ad I.: Ist K eine Kugel in $\bar{\Omega}$ mit einem Mittelpunkt aus Ω, so ist $K \cdot \Omega$ eine Kugel in Ω, also $\bar{\mu}^*(K) = \mu^*(K \cdot \Omega)$ endlich; gehört der Mittelpunkt von K nicht zu Ω, so ist doch K Teilmenge einer solchen Kugel, also $\bar{\mu}^*(K)$ wieder endlich.

Ad V.: Dies braucht nach Anm.[9] a. a. O. nur mit Borelschen O bewiesen zu werden. Wenn Ω absolut Borelsch ist, d. h. Borelsch in seiner vollständigen Hülle $\bar{\Omega}$[11], so schließen wir so: Zu $M \subset \bar{\Omega}$, also $M \cdot \Omega \subset \Omega$, gibt es ein in Ω, also auch in $\bar{\Omega}$ Borelsches $N \supset M \cdot \Omega$ mit $\mu^*(N) = \mu^*(M \cdot \Omega)$. Dann ist $N + (\bar{\Omega} - \Omega)$ auch Borelsch, $\supset M$, und sein $\bar{\mu}^*$ gleich $\mu^*(N) = \mu^*(M \cdot \Omega) = \bar{\mu}^*(M)$, was zu beweisen war.

Wir haben also gezeigt:

Die wesentlichen Sätze 1, 2 und 3 aus der in Anm. [5] zitierten Arbeit gelten auch, wenn der Begriff des m-Raumes weiter gefaßt wird: indem neben der Metrizität und Separabilität keine Vollständigkeit, sondern bloß absolut Borelscher Charakter[11] gefordert wird.

Zu Satz 2 der in Anm.[1] zitierten Arbeit ist dann zu sagen: da die $M(x)$ (in Ω) Borelsch sind, sind sie in diesem Sinne gleichzeitig mit Ω m-Räume ($X^{(0)}$ kann durch seinen maßgleichen F_σ-Kern ersetzt werden, es ist also auch Borelsch, d. h. m-Raum). Im neuen Sinne gilt also Satz 2 ohne weiteres.

2. Im Zusammenhange mit dem Problemkreise des Ergodensatzes sind noch zwei Abhandlungen von Herrn T. Carleman zu nennen, die nach Abfassung der w. o. zitierten Arbeiten des Verf. erschienen sind:

I. Application de la théorie des équations intégrales linéaires aux équations différentielles de la dynamique, Arkiv för Mat., Astr. och Fys., 22 B (1932), No. 7, S. 1–5, eingegangen am 27. Mai 1931, erschienen am 2. März 1932;

II. Application ··· aux systèmes d'équations différentielles non linéaires, Acta Math., 59 (1932), S. 63–87, eingegangen am 10. Februar 1932, erscheint demnächst.

Carleman gelangt in diesen Abhandlungen unabhängig zu Koopmans Operatorenansatz und zu einem, mit demjenigen des Verf. verwandten Beweise des Ergodensatzes[12] unter Beschränkung auf differentiierbare Strömungen.

---

[9] Vgl. Hausdorff, S. 106–107.

[10] Aus Def. 2, a. a. O., Anm. [5].

[11] Vgl. Hausdorff, S. 121 und 208.

[12] Vgl. die Arbeit Koopmans, Proc. Nat. Ac., Bd. 17, Mai 1931, S. 315; diejenige des Verf., Proc. Nat. Ac., Bd. 18, Jan. 1932, S. 70; sowie Teil II, § 2 der in Anm. [1] zitierten.

Da für diese der zur Schar U_t gehörige Operator R (vgl. zur Bezeichnung die Arbeit des Verf. a. a. O. Anm. [1], S. 592, und die zweitvorhergehende, S. 571, sowie a. a. O. Anm. [12], S. 71) direkt angebbar und seine Hypermaximalität leicht beweisbar ist, kann der Stonesche Satz vermieden und unmittelbar die Theorie des Spektrums von R benutzt werden. Diesen Weg schlägt Carleman ein.

Indessen ist sein Beweisverfahren in I. wesentlich lückenhaft, denn seine Betrachtungen beruhen auf der dortigen Formel (4), welche bei den meisten Strömungen unrichtig ist. In der Tat zeigt der Vergleich von (4) mit der vom Verf. benutzten Form der Spektraltheorie, daß das dort auftretende $\theta(p_0, q_0 | \lambda)$ der Integralkern des Operators $1 - E\left(\frac{1}{\lambda}\right)$ $(\lambda > 0)$ bzw. $-E\left(\frac{1}{\lambda}\right)$ $(\lambda < 0)$ sein muß. Bereits auf der Energiefläche des harmonischen linearen Oscillators besitzen aber die genannten Operatoren überhaupt keinen Integralkern. (In Carlemans Ausdrucksweise ist dann $n = 1$, $0 \leqq x_1 \leqq 1$, $x_1 = 0$ mit $x_1 = 1$ zu identifizieren, $A_1 = 1$ — d. h. $\Omega(U) = i\frac{dU}{dx_1}$, falls die Einheiten geeignet gewählt werden. Es liegt also ein reines Punktspektrum bei $\frac{1}{\lambda} = 0, \pm 2\pi, \pm 4\pi, \cdots$ vor, zu 0 gehört die Eigenfunktion 1. Somit wäre $\theta\left(p_0, q_0 \middle| \frac{1}{\pi}\right) - \theta\left(p_0, q_0 | -\frac{1}{\pi}\right) + 1$ der Integralkern von $(1 - E(\pi)) - (-E(-\pi)) + P_{[1]} = 1 + P_{[1]} - (E(\pi) - E(-\pi)) = 1$, aber der Einheitsoperator 1 hat keinen Integralkern. Hierbei haben wir die Normierung $\theta(p_0, q_0 | 0) = 0$ wesentlich benutzt, aber andere Beispiele setzen auch diese nicht voraus.)

In II. wird dieser Fehler durch einen besonderen Kunstgriff (Einführung der Hilfsvariablen ξ, η an Stelle von p_0, q_0) vermieden. Die Rolle der Formel (4) von I. übernimmt die richtige Formel (19) von II., und der Beweis wird lückenlos.

YALE UNIVERSITY
LIBRARY

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

EDITED BY

EINAR HILLE S. LEFSCHETZ

WITH THE COÖPERATION OF

ORMOND STONE
J. W. ALEXANDER
H. BATEMAN
G. D. BIRKHOFF
L. P. EISENHART
OYSTEIN ORE
J. F. RITT
J. D. TAMARKIN
H. S. VANDIVER
OSWALD VEBLEN
J. H. M. WEDDERBURN
A. PELL-WHEELER
NORBERT WIENER

PUBLISHED BY THE

PRINCETON UNIVERSITY PRESS

SECOND SERIES, VOL 33, NO. 1

JANUARY 1932

PRINCETON, N. J.

1932

According to an agreement between the Mathematical Association of America and the editors of the Annals of Mathematics, the Association contributes to the support of the Annals, and the Annals is supplied to individual members of the Association at one half of the regular price. In consequence of this agreement the volume of the Annals was increased by 100 pages to be devoted to expository and historical articles in so far as suitable articles of this class are obtainable.

A number of the expository articles which have already been published are available in separate form and are listed for sale on the inside of the back cover of this number of the Annals. The regular subscription price of the Annals is $5.00 a volume.

Copies of the following memoirs can be obtained by addressing THE PRINCETON UNIVERSITY PRESS, Princeton, N. J.:

An elementary exposition of the theory of the gamma function. By J. L. W. V. JENSEN. Authorized translation with additional notes by T. H. GRONWALL. 43 pages. Price 50 cents.

Fermat's last theorem and the origin and nature of the theory of algebraic numbers. By L. E. DICKSON. 27 pages. Price 35 cents.

Factorization of analytic functions of several variables. By W. F. OSGOOD. 19 pages. Price 25 cents.

Investigation of a class of fundamental inequalities in the theory of analytic functions. By J. L. W. V. JENSEN. Authorized translation from the Danish by T. H. GRONWALL. 29 pages. Price 40 cents.

An introduction to the theory of elliptic functions. By GÖSTA MITTAG-LEFFLER. Authorized translation by EINAR HILLE. 81 pages. Price 90 cents.

Differential equations from the group standpoint. By L. E. DICKSON. 92 pages. Price $1.00.

On various conceptions of correlation. By F. M. WEIDA. 37 pages. Price 75 cents.

Contact transformations. By L. P. EISENHART. 40 pages. Price 80 cents.

A survey of the theory of small samples. By P. R. RIDER. 52 pages. Price $1.00.

CONTENTS

	Page
Tauberian theorems. By N. WIENER	1
Summation of series and conformal mapping. By T. H. GRONWALL	101
A rule of signs involving certain orthogonal polynomials. By M. MARDEN	118
On a certain inequality of Steiner. By E. J. MCSHANE	125
The expansion of analytic functions in generalized Lambert series. By J. M. FELD	139
Arithmetical periodicities of Bessel functions. By D. H. LEHMER	143
A note on hypergeometric functions of two variables. By T. H. RAWLES	151
Sur une propriété métrique de la dimension. Par L. PONTRJAGIN et L. SCHNIRELMANN	156
Über stetige algebraische Körper. Von L. PONTRJAGIN	163
Zur Begründung der projektiven Geometrie. Von A. KOLMOGOROFF	175
On the resultant of three double binary forms. By N. H. MCCOY	177
On the composition of dependent elementary errors. By C. C. CRAIG	184

ANNALS OF MATHEMATICS

Published at Princeton, N.J. Subscription price, $5 a volume (four numbers) in advance. Single copies $1.50. Subscriptions, orders for back numbers, and changes of address should be sent to the PRINCETON UNIVERSITY PRESS, PRINCETON, NEW JERSEY.

Manuscripts and all editorial correspondence should be addressed to THE ANNALS OF MATHEMATICS, Fine Hall, PRINCETON, NEW JERSEY. Manuscripts should be typewritten, with the exception of formulæ, and must be in final form with all references filled in.

Authors receive gratis 100 reprints of each article, postage prepaid. Additional copies will be furnished at cost.

PRINTED BY LÜTCKE & WULFF, HAMBURG, GERMANY

YALE UNIVERSITY
MAY 14 1932
LIBRARY

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

EDITED BY

EINAR HILLE S. LEFSCHETZ

WITH THE COÖPERATION OF THE

DEPARTMENT OF MATHEMATICS OF PRINCETON UNIVERSITY

AND OF

ORMOND STONE
H. BATEMAN
G. D. BIRKHOFF
OYSTEIN ORE
J. F. RITT
J. D. TAMARKIN
H. S. VANDIVER
A. PELL-WHEELER
NORBERT WIENER

PUBLISHED BY THE

PRINCETON UNIVERSITY PRESS

SECOND SERIES, VOL. 33, NO. 2

APRIL 1932

PRINCETON, N. J.

1932

According to an agreement between the Mathematical Association of America and the editors of the Annals of Mathematics, the Association contributes to the support of the Annals, and the Annals is supplied to individual members of the Association at one half of the regular price. In consequence of this agreement the volume of the Annals was increased by 100 pages to be devoted to expository and historical articles in so far as suitable articles of this class are obtainable.

A number of the expository articles which have already been published are available in separate form and are listed for sale on the inside of the back cover of this number of the Annals. The regular subscription price of the Annals is $5.00 a volume.

Copies of the following memoirs can be obtained by addressing THE PRINCETON UNIVERSITY PRESS, Princeton, N. J.:

An elementary exposition of the theory of the gamma function. By J. L. W. V. JENSEN. Authorized translation with additional notes by T. H. GRONWALL. 43 pages. Price 50 cents.

Fermat's last theorem and the origin and nature of the theory of algebraic numbers. By L. E. DICKSON. 27 pages. Price 35 cents.

Factorization of analytic functions of several variables. By W. F. OSGOOD. 19 pages. Price 25 cents.

Investigation of a class of fundamental inequalities in the theory of analytic functions. By J. L. W. V. JENSEN. Authorized translation from the Danish by T. H. GRONWALL. 29 pages. Price 40 cents.

An introduction to the theory of elliptic functions. By GÖSTA MITTAG-LEFFLER. Authorized translation by EINAR HILLE. 81 pages. Price 90 cents.

Differential equations from the group standpoint. By L. E. DICKSON. 92 pages. Price $1.00.

On various conceptions of correlation. By F. M. WEIDA. 37 pages. Price 75 cents.

Contact transformations. By L. P. EISENHART. 40 pages. Price 80 cents.

A survey of the theory of small samples. By P. R. RIDER. 52 pages. Price $1.00.

Tauberian theorems. By NORBERT WIENER. 100 pages. Price $1.00.

CONTENTS

	Page
On the real multiplication of elliptic functions. By E. T. Bell	207
Elementary transformations. By E. R. Lorch	214
A problem in the simultaneous reduction of two quadratic forms in infinitely many variables. By R. L. Anderson	229
The problem of Bolza in the calculus of variations. By G. A. Bliss	261
An inequality for the Bessel functions of the first kind with imaginary argument. By T. H. Gronwall	275
A special conformally euclidean space of three dimensions occurring in wave mechanics. By T. H. Gronwall	279
Über adjungierte Funktionaloperatoren. Von J. v. Neumann	294
Algebras of degree 2^e and pure Riemann matrices. By A. A. Albert	311
A few desirable modifications in the literature of group theory. By G. A. Miller	319
The linear form of numbers represented by a homogeneous polynomial in any number of variables. By M. Ward	324
A proof of a certain congruence. By W. L. G. Williams	327
A set of postulates for plane geometry, based on scale and protractor. By G. D. Birkhoff	329
A set of postulates for the foundation of logic. By A. Church	346
Circuits upon polyhedra. By C. N. Reynolds, Jr.	367
Group invariants and torsion coefficients. By A. B. Brown	373

ANNALS OF MATHEMATICS

Published at Princeton, N.J. Subscription price, $5 a volume (four numbers) in advance. Single copies $1.50. Subscriptions, orders for back numbers, and changes of address should be sent to the Princeton University Press, Princeton, New Jersey.

Manuscripts and all editorial correspondence should be addressed to The Annals of Mathematics, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten, with the exception of formulæ, and must be in final form with all references filled in.

Authors receive gratis 100 reprints of each article, postage prepaid. Additional copies will be furnished at cost.

Printed by Lütcke & Wulff, Hamburg, Germany

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

EDITED BY

EINAR HILLE S. LEFSCHETZ

WITH THE COÖPERATION OF THE

DEPARTMENT OF MATHEMATICS OF PRINCETON UNIVERSITY

AND OF

ORMOND STONE J. F. RITT
H. BATEMAN J. D. TAMARKIN
G. D. BIRKHOFF H. S. VANDIVER
OYSTEIN ORE A. PELL-WHEELER
NORBERT WIENER

PUBLISHED BY THE

PRINCETON UNIVERSITY PRESS

SECOND SERIES, VOL. 33, NO. 3

JULY 1932

PRINCETON, N. J.

1932

YALE UNIVERSITY
AUG 13 1932
LIBRARY.

According to an agreement between the Mathematical Association of America and the editors of the Annals of Mathematics, the Association contributes to the support of the Annals, and the Annals is supplied to individual members of the Association at one half of the regular price. In consequence of this agreement the volume of the Annals was increased by 100 pages to be devoted to expository and historical articles in so far as suitable articles of this class are obtainable.

A number of the expository articles which have already been published are available in separate form and are listed for sale on the inside of the back cover of this number of the Annals. The regular subscription price of the Annals is $5.00 a volume.

Copies of the following memoirs can be obtained by addressing THE PRINCETON UNIVERSITY PRESS, Princeton, N. J.:

An elementary exposition of the theory of the gamma function. By J. L. W. V. JENSEN. Authorized translation with additional notes by T. H. GRONWALL. 43 pages. Price 50 cents.

Fermat's last theorem and the origin and nature of the theory of algebraic numbers. By L. E. DICKSON. 27 pages. Price 35 cents.

Factorization of analytic functions of several variables. By W. F. OSGOOD. 19 pages. Price 25 cents.

Investigation of a class of fundamental inequalities in the theory of analytic functions. By J. L. W. V. JENSEN. Authorized translation from the Danish by T. H. GRONWALL. 29 pages. Price 40 cents.

An introduction to the theory of elliptic functions. By GÖSTA MITTAG-LEFFLER. Authorized translation by EINAR HILLE. 81 pages. Price 90 cents.

Differential equations from the group standpoint. By L. E. DICKSON. 92 pages. Price $1.00.

On various conceptions of correlation. By F. M. WEIDA. 37 pages. Price 75 cents.

Contact transformations. By L. P. EISENHART. 40 pages. Price 80 cents.

A survey of the theory of small samples. By P. R. RIDER. 52 pages. Price $1.00.

Tauberian theorems. By NORBERT WIENER. 100 pages. Price $1.00.

CONTENTS

	Page
On normal products of algebras. By J. Levitzki	377
Definition of a field by four postulates. By W. A. Hurwitz	403
Note on multiple Dirichlet and multiple factorial series. By C. R. Adams	406
On deferred Cesàro means. By R. P. Agnew	413
Extension of the notion of the limit of the sum of terms of an infinite series. By G. F. Woronoi	422
A remark on Nörlund's method of summation. By R. E. Gilman	429
A class of functions continuous but not absolutely continuous. By R. E. Gilman	433
Relative summability. By R. L. Jeffery	443
On the semi-continuity of double integrals in the calculus of variations. By E. J. McShane	460
Some applications of the calculus of variations to Riemannian geometry. By I. J. Schoenberg	485
Groups of motions in spaces admitting absolute parallelism. By H. P. Robertson	496
On duality and intersection chains in combinatorial analysis situs. By J. H. C. Whitehead	521
On separable spaces. By S. Lefschetz	525
A classification of the homology groups of compact spaces. By J. W. Alexander and L. W. Cohen	538
Über einen Satz von Herrn M. H. Stone. Von J. v. Neumann	567
Einige Sätze über meßbare Abbildungen. Von J. v. Neumann	574
Zur Operatorenmethode in der klassischen Mechanik. Von J. v. Neumann	587
On one-parameter unitary groups in Hilbert space. By M. H. Stone	643

ANNALS OF MATHEMATICS

Published at Princeton, N.J. Subscription price, $5 a volume (four numbers) in advance. Single copies $1.50. Subscriptions, orders for back numbers, and changes of address should be sent to the Princeton University Press, Princeton, New Jersey.

Manuscripts and all editorial correspondence should be addressed to The Annals of Mathematics, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten, with the exception of formulæ, and must be in final form with all references filled in.

Authors receive gratis 100 reprints of each article, postage prepaid. Additional copies will be furnished at cost.

Printed by Lütcke & Wulff, Hamburg, Germany

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

EDITED BY

EINAR HILLE S. LEFSCHETZ

WITH THE COÖPERATION OF THE

DEPARTMENT OF MATHEMATICS OF PRINCETON UNIVERSITY

AND OF

ORMOND STONE
H. BATEMAN
G. D. BIRKHOFF
OYSTEIN ORE
J. F. RITT
J. D. TAMARKIN
H. S. VANDIVER
A. PELL-WHEELER
NORBERT WIENER

PUBLISHED BY THE

PRINCETON UNIVERSITY PRESS

SECOND SERIES, VOL. 33, NO. 4

OCTOBER 1932

YALE UNIVERSITY
NOV 7 1932
LIBRARY

PRINCETON, N. J.

1932

According to an agreement between the Mathematical Association of America and the editors of the Annals of Mathematics, the Association contributes to the support of the Annals, and the Annals is supplied to individual members of the Association at one half of the regular price. In consequence of this agreement the volume of the Annals was increased by 100 pages to be devoted to expository and historical articles in so far as suitable articles of this class are obtainable.

A number of the expository articles which have already been published are available in separate form and are listed for sale on the inside of the back cover of this number of the Annals. The regular subscription price of the Annals is $5.00 a volume.

Copies of the following memoirs can be obtained by addressing THE PRINCETON UNIVERSITY PRESS, Princeton, N. J.:

An elementary exposition of the theory of the gamma function. By J. L. W. V. JENSEN. Authorized translation with additional notes by T. H. GRONWALL. 43 pages. Price 50 cents.

Fermat's last theorem and the origin and nature of the theory of algebraic numbers. By L. E. DICKSON. 27 pages. Price 35 cents.

Factorization of analytic functions of several variables. By W. F. OSGOOD. 19 pages. Price 25 cents.

Investigation of a class of fundamental inequalities in the theory of analytic functions. By J. L. W. V. JENSEN. Authorized translation from the Danish by T. H. GRONWALL. 29 pages. Price 40 cents.

An introduction to the theory of elliptic functions. By GÖSTA MITTAG-LEFFLER. Authorized translation by EINAR HILLE. 81 pages. Price 90 cents.

Differential equations from the group standpoint. By L. E. DICKSON. 92 pages. Price $1.00.

On various conceptions of correlation. By F. M. WEIDA. 37 pages. Price 75 cents.

Contact transformations. By L. P. EISENHART. 40 pages. Price 80 cents.

A survey of the theory of small samples. By P. R. RIDER. 52 pages. Price $1.00.

Tauberian theorems. By NORBERT WIENER. 100 pages. Price $1.00.

CONTENTS

	Page
Pairs of curves in an S_n. By G. A. PARKINSON	649
The area and boundary of minimal surfaces. By E. F. BECKENBACH	658
Equivalent continuous groups. By L. P. EISENHART	665
On the cellular subdivision of n-dimensional regions. By S. S. CAIRNS	671
Locally homogeneous spaces in differential geometry. By J. H. C. WHITEHEAD	681
The coloring of graphs. By H. WHITNEY	688
Geodesics on a two-dimensional Riemannian manifold with periodic coefficients. By G. A. HEDLUND	719
On the singular points in the problem of inversion. By W. F. OSGOOD	740
On the Weierstrass condition for the problem of Bolza in the calculus of variations. By L. M. GRAVES	747
On a theorem of Gross and Iversen. By J. L. DOOB	753
On the summability of power series. By L. S. BOSANQUET	758
Non-summability of the conjugate series of a Fourier series. By B. N. PRASAD	771
On a method of summation of Fourier series. By A. F. MOURSUND	773
Remarks on a problem of Toeplitz. By F. BOHNENBLUST and E. HILLE	785
A note on Tauberian theorems. By N. WIENER	787
Note on a preceding paper. By G. D. BIRKHOFF	788
Zusätze zur Arbeit „Zur Operatorenmethode ...". Von J. v. NEUMANN	789

ANNALS OF MATHEMATICS

Published at Princeton, N.J. Subscription price, $5 a volume (four numbers) in advance. Single copies $1.50. Subscriptions, orders for back numbers, and changes of address should be sent to the PRINCETON UNIVERSITY PRESS, PRINCETON, NEW JERSEY.

Manuscripts and all editorial correspondence should be addressed to THE ANNALS OF MATHEMATICS, Fine Hall, PRINCETON, NEW JERSEY. Manuscripts should be typewritten, with the exception of formulæ, and must be in final form with all references filled in.

Authors receive gratis 100 reprints of each article, postage prepaid. Additional copies will be furnished at cost.

PRINTED BY LUTCKE & WULFF, HAMBURG, GERMANY